Application of Robust Statistics to Asset Allocation Models

by
Xinfeng Zhou

M.S., Chemistry
University of Washington, 2001

Submitted to Sloan School of Management
in partial fulfillment of the requirements for the degree of
Master of Science in Operations Research
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2006

Signature of Author ....
Sloan School of Management
Operations Research Center
May 2006

Certified by ....
Roy E. Welsch
Professor of Management Science, Statistics, and Engineering Systems
Thesis Supervisor

Certified by ....
George Verghese
Professor of Electrical Engineering
Thesis Supervisor

Accepted by ....
James B. Orlin
Edward Pennell Brooks Professor of Operations Research
Codirector, Operations Research Center

Application of Robust Statistics to Asset Allocation Models

By
Xinfeng Zhou

Submitted to Sloan School of Management
On May 17, 2006 in partial fulfillment of the requirements
for the degree of Master of Science in Operations Research

ABSTRACT

Many strategies for asset allocation involve the computation of expected returns and the covariance or correlation matrix of financial instruments returns. How much of each instrument to own is determined by an attempt to minimize risk (the variance of linear combinations of investments in these financial assets) subject to various constraints such as a given level of return, concentration limits, etc. The expected returns and the covariance matrix contain many parameters to estimate and two main problems arise. First, the data will very likely have outliers that will seriously affect the covariance matrix. Second, with so many parameters to estimate, a large number of observations are required and the nature of markets may change substantially over such a long period. In this thesis we use robust covariance procedures, such as FAST-MCD, quadrant-correlation-based covariance and 2D-Huber-based covariance, to address the first problem and regularization (Bayesian) methods that fully utilize the market weights of all assets for the second. High breakdown affine equivariant robust methods are effective, but tend to be costly when cross-validation is required to determine regularization parameters. We, therefore, also consider non-affine invariant robust covariance estimation. When back-tested on market data, these methods appear to be effective in improving portfolio performance. In conclusion, robust asset allocation methods have great potential to improve risk-adjusted portfolio returns and therefore deserve further exploration in investment management research.

Thesis Supervisor: Roy E. Welsch
Title: Professor of Management Science, Statistics, and Engineering Systems

Thesis Supervisor: George Verghese
Title: Professor of Electrical Engineering

Acknowledgements

I would like to express my greatest gratitude to my advisors, Professor Roy Welsch and Professor George Verghese for their constant guidance, advice and encouragement. Working with Professor Welsch was a really rewarding experience. He truly shaped my understanding of statistics. Professor Verghese has been an inspiring mentor over the years who continuously encouraged and supported my efforts in knowledge-building and research. I am also deeply indebted to my PhD advisor, Professor Peter Dedon from the department of Biological Engineering Division, for his understanding and support over the last five years. I also want to thank Sameer Jain for his expertise, support and inputs into my work. Last but not at least, I would like to express many thanks to Dr. Jan-Hein Cremers at State Street Corporation for providing me opportunities to work on asset management projects and for inspiring my interests in topics covered in this thesis.

This thesis is dedicated to the memory of my sister, Xinli Zhou.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction of Asset Allocation Models

1.1 Overview

Asset allocation is the process that investors use to determine the asset classes in which to invest in and the weight for each asset class. Past studies have shown that asset allocation explains 75 - 90% of the return variation and is the single most important factor determining the variability of portfolio performance. The objective of an asset allocation model is to find the right asset mix that provides the appropriate combination of expected return and risk that allows investors to achieve their financial goals. Harry Markowitz's mean-variance portfolio theory is by far the most well-known and well-studied asset allocation model for both academic researchers and practitioners alike (*1, 2*). The crux of mean-variance portfolio theory assumes that investors prefer (1) higher expected returns for a given level of standard deviation/variance and (2) lower standard deviations/variances for a given level of expected return. Portfolios that provide the maximum expected return for a given standard deviation and the minimum standard deviation for a given expected return are termed efficient portfolios and those that don't are termed inefficient portfolios.

Although intuitively and theoretically appealing, the application of mean-variance portfolio optimization has been hindered by the difficulty in accurately estimating model inputs, the expected returns and the covariance matrix of the assets. The goal of this thesis is to address this critical problem from different perspectives, with an emphasis on robust statistics and Bayesian approaches. In doing so, the thesis is organized as five Chapters:

In Chapter 1, we first present the economic and mathematical background of the mean-variance portfolio optimization model and discuss the importance as well as the difficulty in estimating the model inputs, the expected return and the covariance matrix.

In Chapter 2, we investigate and apply some of the existing factor models and Bayesian shrinkage models as well as Generalized Autoregressive Conditional Heteroskedastic (GARCH) models to estimate the expected returns and the covariance matrix. Using the performance results of a US industrial selection portfolio, we show that "optimal" portfolios selected by shrinkage models have limited success when the number of assets N is of the same order as the number of return observations T. The GARCH models, which automatically include all historical information by exponential weighting, yield much better results.

In Chapter 3, we investigate and expand some of the robust statistical approaches to estimate the expected returns and the covariance matrix. Beside traditional robust methods, such as the least absolute deviation (LAD) method and the Spearman rank correlation, we focus more on recent developments such as the quadrant-correlation-based covariance and the 2D-Huber-based covariance to reduce or eliminate the effect of outliers. These new models prove to be more valuable than shrinkage models and GARCH models in dramatically improving risk-adjusted portfolio performance and reducing asset turnovers.

In Chapter 4, we investigate some of the more complex Bayesian approaches with emphasis on regularization models. We show that regularization models yield portfolios with significant increases in risk-adjusted portfolio performance, especially when transaction costs are taken into consideration. Our results also indicate that L_1 regularization methods may yield better results than L_2 regularization methods. Overall L_1 regularization methods also outperform the robust covariance estimation methods (*e.g.*, 2D-Huber), however the improvement is achieved at the cost of higher computational complexity.

Finally, in Chapter 5 we conclude this thesis by summarizing our results and offering possible directions for future research.

1.2 Foundation of Mean-Variance Portfolio Optimization

Mean-variance portfolio theory is built upon a basic economic principle, utility-maximization under economic constraints. In economics, utility is a measure of the happiness or satisfaction gained by consuming goods and services. Financial economists believe that rational investors make investment decisions (as well as consumption decisions) to maximize their lifetime expected utility under the budget constraints. The discrete version of the problem with T periods can be mathematically formulated as:

$$Max\ E_0[U(C_1, C_2, \cdots, W_T)]$$

$$\begin{aligned} s.t.\quad & W_t + Y_t - C_t = I_t = \sum_{i=1}^{N} n_{it} P_{it} \quad \text{(budget constraint)} \\ & W_{t+1} = I_t \sum_{i=1}^{N} w_{it} Z_{it} = (W_t + Y_t - C_t) \times \left[\sum_{i=1}^{n} \varpi_{it} Z_{it} \right] \quad \text{(wealth dynamic)} \end{aligned} \tag{1.1}$$

where W_t, W_{t+1} are the wealth of the investor at time t and $t+1$, respectively; Y_t is the (labor) income of the investor at period t; C_t is the consumption of the investor at time t; I_t is the investment made at time t; N is the number of assets; P_{it} is the price of the ith asset at time t; n_{it}

is the number of shares invested in the ith asset; $w_{it} \equiv n_{it} P_{it} / I_t$ is the weight of the ith asset in the portfolio, $\sum_{i=1}^{n} w_{it} = 1$; $Z_{it} \equiv P_{it+1} / P_{it}$ is the gross return of the ith asset.

The budget constraint equation states that the amount an investor can put into a portfolio is the amount of wealth at time t plus income minus consumption during the period. The wealth dynamic equation shows that the portfolio return equals the weighted average returns of individual assets in the investment.

The objective of this optimization problem is to maximize expected utilities over the lifetime of an investor. The utility function is a function of consumption and final wealth, which assigns a happiness score to each consumption set and final wealth. Different investors may have different utility functions; nevertheless the utility function of a rational investor should satisfy the following four basic properties:

1. Completeness: either $U_1 > U_2$ or $U_1 < U_2$ or $U_1 = U_2$;
2. Transitivity: $U_1 > U_2, U_2 > U_3 \Rightarrow U_1 > U_3$;
3. Non-satiation: more wealth/consumption is better than less wealth/consumption;
4. Law of diminishing returns: diminishing marginal utility of wealth/consumption.

Although the mathematical formulation is simple, the problem is a gigantic dynamic programming dilemma that can never be solved because all Y_t, P_{it}, Z_{it} and even C_t (which are often influenced by inflation) are random numbers. Even for a small number of periods, the dimension of the problem becomes intractable.

Instead, a much simpler static version of the problem that only considers two time periods, *0* and *1*, is often used in practice. The problem considers an investor with current wealth W_0 that needs to be invested in *n* assets, which will yield future wealth W_1. The utility at period *1* is determined by W_1 and the simplified investment problem can be formulated as

$$\underset{\{w_i\}}{Max}\ E[U(W_1)]$$

$$\begin{aligned} s.t.\quad & W_1 = W_0 \sum_{i=1}^{N} w_i(1+r_i) \\ & \sum_{i=1}^{N} w_i = 1 \end{aligned} \qquad (1.2)$$

where $r_i \equiv P_{i1}/P_{i0} - 1$ is the net return of the *i*th asset; w_i is the weight of the *i*th asset in the portfolio. Again the weights of all assets add up to one, and the portfolio return is the weighted average return of all the assets.

In this version of the problem, the utility function is a function of wealth level *W,* which reflects an investor's preferences and also reveals his/her attitude toward risk. Non-satiation indicates that investors prefer more wealth to less wealth, which means positive marginal utility: $\frac{dU}{dW} > 0$ $\forall W$. The Law of diminishing returns indicates that the utility function is concave, $\frac{d^2U}{dW^2} < 0$ $\forall W$. Power utility functions $U = f(W) = \frac{1}{\gamma} \times W^{\gamma}$ are commonly used by financial economists. A log wealth utility function is a special case of the power utility. As γ approaches 0, utility approaches the natural logarithm of wealth. A γ equal to 1/2 implies less risk aversion than log wealth, while a γ equal to -1 implies greater risk aversion (*3*). Applying the second-order Taylor series to any utility functions, we can express the utility as

$$U(W_1) = U(\bar{W}_1) + U'(\bar{W})(W_1 - \bar{W}_1) + \tfrac{1}{2!}U''(\bar{W}_1)(W_1 - \bar{W}_1)^2 + \tfrac{1}{3!}U^3(W^*)(W_1 - \bar{W}_1)^3 \qquad (1.3)$$

where $\bar{W}_1$ is the expected wealth at time 1; $U'(\bar{W})$ and $U''(\bar{W}_1)$ are the first and the second derivatives of utility function at $\bar{W}$; $U^3(W^*)$ is the third derivative of the utility function at some W^* between W_1 and $\bar{W}_1$.

If we approximate the function by ignoring the error term $\tfrac{1}{3!}U^3(W^*)(W_1 - \bar{W}_1)^3$, $E[U(W_1)]$ can be written as $E[U(W_1)] \approx U(\bar{W}_1) + U'(\bar{W})E[(W_1 - \bar{W}_1)] + \tfrac{1}{2!}U''(\bar{W}_1)E[(W_1 - \bar{W}_1)^2] = U(\bar{W}_1) + \tfrac{1}{2!}U''(\bar{W}_1)\sigma_W^2$. This equation shows that if the expected wealth $\bar{W}_1$ is held constant, the lower the variance the higher the expected utility since $U''(\bar{W}_1) < 0$. If the variance is held constant, the higher the expected wealth $\bar{W}$ the higher the expected utility since $U'(\bar{W}_1) > 0$.

Since $U(\bar{W}_1)$ is a monotone increasing function of $\bar{W}_1$ and $\bar{W}_1 = W_0 \times (1 + E[r_p])$ is a monotone increasing function of $E[r_p]$, $U(\bar{W}_1)$ is also a monotone increasing function of $E[r_p]$. The approximation indicates that investors prefer higher expected portfolio returns and lower variance. We can represent the approximation using a quadratic utility function:

$$U = E[r_p] - \tfrac{1}{2}\lambda \operatorname{var}(r_p) \qquad (1.4)$$

where $E[r_p]$ is the expected portfolio return; $\operatorname{var}(r_p)$ is the variance of the portfolio return; λ is the risk aversion coefficient ($\lambda > 0$) with a larger value of λ indicating higher risk aversion.

If the portfolio return follows a normal distribution, then $E\left[(W_1 - \bar{W}_1)^3\right] = 0$ and the expected utility function is quadratic without approximation. So one underlying assumption behind mean/variance portfolio theory is that the portfolio return is normally distributed or investors have a quadratic utility function. It is worth noting that neither condition is met in reality. Simple quadratic functions may not be consistent with the fundamental property of a utility function since it indicates that, at certain wealth levels, the function has negative marginal utility and the investors prefer less wealth to more wealth (satiation). Asset returns often have fat tails and/or are positive/negatively skewed. So mean-variance optimization, as many other mathematical models, is a simplified representation of reality.

1.3 The Mean-Variance Efficient Frontier

For a portfolio with N risky assets to invest in, the portfolio return is the weighted average return of each asset: $r_p \equiv w_1 r_1 + w_2 r_2 + \cdots + w_N r_N = w' r$. So the expected return and the variance of portfolio can be expressed as

$$\mu_p = w_1\mu_1 + w_2\mu_2 + \cdots + w_N\mu_N = w'\mu$$
$$\mathrm{var}(r_p) \equiv \mathrm{var}(w_1 r_1 + w_2 r_2 + \cdots + w_N r_N) = \sum_{i=1}^{N} \sigma_i^2 w_i^2 + \sum_{i \neq j} \sigma_{ij} w_i w_j = w' \Sigma w \qquad (1.5)$$

where $w_i, \forall i = 1, \cdots, N$ is the weight of the ith asset in the portfolio; $r_i, \forall i = 1, \cdots, N$ is the return of the ith asset in the portfolio, $r_i \equiv P_{i1} / P_{i0} - 1$; $\mu_i, \forall i = 1, \cdots, N$ is the expected return of the ith asset in the portfolio; w is a $N \times 1$ column vector of w_i s; r is a $N \times 1$ column vector of r_i s; μ

is a $N \times 1$ column vector of μ_i s; Σ is the covariance matrix of the returns of N assets, an $N \times N$ matrix.

Since the optimal portfolio minimizes the variance of return for a given level of expected return, we can formulate the following problem to assign optimal weight to each asset and identify the efficient portfolio:

$$\begin{aligned} &\min_{\varpi} \tfrac{1}{2} w'\Sigma w \\ &s.t.\ w'\mu = \mu_p,\ w'e = 1 \end{aligned} \tag{1.6}$$

where e is $N \times 1$ column vector with all elements 1.

This problem can be solved in closed form using the method of Lagrange (*4*):

$L = \frac{1}{2} w'\Sigma w + \gamma(\mu_p - w'\mu) + \lambda(1 - w'e)$ where λ is a $1 \times k$ vector.

To minimize L, take the derivative of w, γ and λ

$$\left.\begin{aligned} \frac{\partial \mathrm{L}}{\partial w} = 0 \quad &\Leftrightarrow \quad \Sigma w - \gamma\mu - \lambda e = 0 \\ \frac{\partial \mathrm{L}}{\partial \gamma} = 0 \quad &\Leftrightarrow \quad \mu_p - w'\mu = 0 \\ \frac{\partial \mathrm{L}}{\partial \lambda} = 0 \quad &\Leftrightarrow \quad 1 - w'e = 0 \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} &w^* = \lambda\Sigma^{-1}e + \gamma\Sigma^{-1}\mu \quad \text{(Minimum varinace portfolio)} \\ &\lambda = \frac{C - \mu_p B}{D} \\ &\gamma = \frac{\mu_p A - B}{D} \end{aligned}\right. \tag{1.7}$$

where $A = e'\Sigma^{-1}e > 0,\ B = e'\Sigma^{-1}\mu,\ C = \mu'\Sigma^{-1}\mu > 0,\ D = AC - B^2$.

For each specified μ_p, we can easily derive the optimal portfolio weights w^* and the corresponding minimum variance $\sigma_p^2 = w^{*\prime}\Sigma w^*$. Combining the results, we have a hyperbola

$\sigma_p^2 = \lambda + \gamma\mu_p = A\mu_p^2 - 2B\mu_p + C / D$, which describes the relationship between the expected return and the standard deviation.

To minimize the variance, set $d(\sigma_p^2)/d\mu_p = 0$, and we have the global minimum variance portfolio:

$$\mu_g = B/A \Rightarrow \sigma_g^2 = 1/A, w_g = \Sigma^{-1}e/A. \tag{1.8}$$

As the result clearly shows, the **global minimum variance portfolio (GMVP)** does not require an estimation of expected returns μ, which makes it a perfect choice to independently test covariance matrix estimation.

If a riskless asset (an asset with zero variance) with return r_f is introduced, the problem becomes:

$$\begin{aligned} &\min_w \tfrac{1}{2} w'\Sigma w \\ &s.t. \;\; w'\mu + (1 - w'e) r_f = \mu_p \end{aligned} \tag{1.9}$$

Again using the Lagrange method, we have

$$w^* = \lambda \Sigma^{-1}(\mu - r_f \text{ e}) \text{ and } \lambda = \frac{\mu_p - r_f}{(\mu - r_f e)^T \Sigma^{-1}(\mu - r_f e)} = \frac{\mu_p - r_f}{C - 2r_f B + r_f^2 A} \tag{1.10}$$

The relationship between expected portfolio return and the variance of efficient portfolio can be expressed as:

$$\sigma_p^2 = (\mu_p - r_f)^2 / E, E = C - 2r_f B + r_f^2 A = (\mu - r_f e)'\Sigma^{-1}(\mu - r_f e) \tag{1.11}$$

The equation is more frequently expressed as $\textit{Sharpe ratio} = \frac{\mu_p - r_f}{\sigma_p} = \sqrt{E}$, which is the foundation of the well-known Capital Asset Pricing Model (CAPM) introduced by Sharpe and Lintner (*5, 6*).

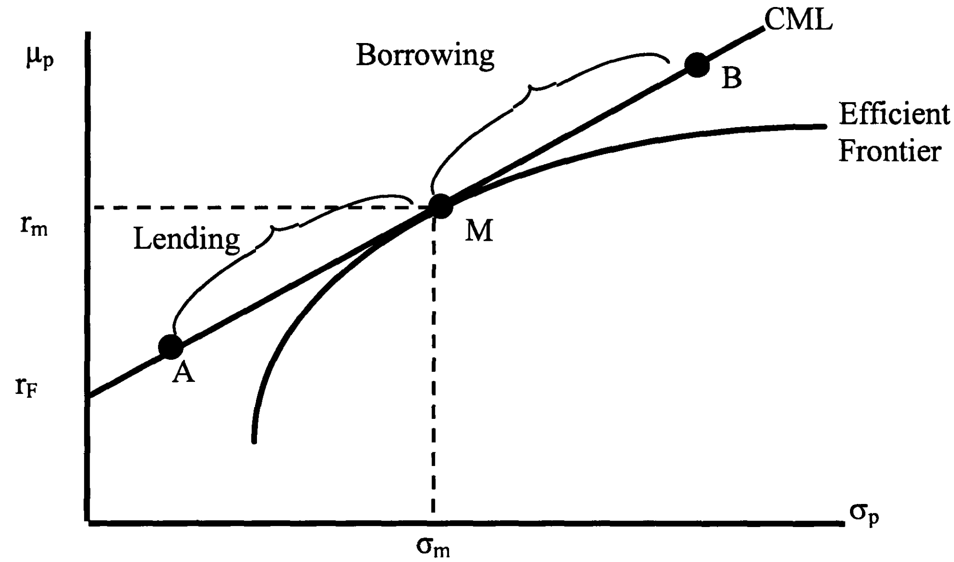

Figure 1.1. Efficient frontier and Capital Market Line

As shown in Figure 1.1, the relationship between the expected return and the risk (as measured by standard deviation) is a linear relationship. The addition of a riskless asset changes the hyperbolic efficient frontier to a straight line, which is tangent to the efficient frontier with only risky assets. It shows that investors should only invest in a combination of the risk free asset and the tangent portfolio. Under a set of strict assumptions, such as perfect rationality and homogenous expectations of investors, Sharpe and Litner showed that the tangent portfolio must be the market portfolio (a value weighted index of the entire market) and the tangent line is called the capital market line (CML). The mathematics behind this is complex, but the intuition is straightforward. In equilibrium, all stocks must be held (markets have to clear). Therefore, prices will adjust to make sure that holding the market portfolio is the best choice. This result is often referred to as CAPM since it provides a prediction of the relationship between an asset's risk and its expected return. Based on CAPM, the expected excess return of any asset can be expressed as a function of market return:

$$\mu_i - r_f = \beta_i(E[r_m] - r_f) \quad (1.12)$$

where $\beta_i = \dfrac{\text{cov}(r_i, r_m)}{\text{var}(r_m)} = \rho_{i,m} \times \dfrac{\sigma_i}{\sigma_m}$ is the sensitivity of the asset returns to the market returns; $E[r_m]$ is the expected return on the market portfolio.

CAPM decomposes an asset's risk into systematic risk and idiosyncratic risk. Systematic risk is determined by the covariance of the asset with the market portfolio, and the rest of the risk is idiosyncratic risk. The gist of CAPM is that only systematic risk will be compensated with higher expected return, while idiosyncratic risk is not compensated since it can be diversified away. US treasury bills are often considered to be a risk-free asset for US investors. The market portfolio is supposed to include all risky assets. Yet in practice, S&P 500 and other traditional market-weighted indices are often used as a proxy for the market portfolio.

Legal restrictions, investment policies and investors' attitudes often impose constraints on the assets classes to invest in as well as portfolio weights. One of the most common constraints is to exclude short sales ($w_i \geq 0, i = 1, \cdots, N$), which is the legal requirement of many mutual funds and pension funds. Taking different constraints into consideration, a more general portfolio optimization problem can be expressed as:

$$\begin{aligned} &\min_{\varpi} \tfrac{1}{2} w' \Sigma w \\ &s.t. \;\; w' \mu = \mu_p \\ &\qquad Aw \geq c \\ &\qquad Bw = d \end{aligned} \quad (1.13)$$

where $Aw \geq c$ represents the inequality constraints of asset weights and $Bw = d$ represents the equality constraints.

The problem is a quadratic minimization problem with linear constraints. The necessary and sufficient conditions for the problem are given by the Kuhn-Tucker conditions. All these optimization problems can be solved efficiently using quadratic programming algorithms, such as the interior point method.

1.4 Advantages and Disadvantages of Mean-Variance Optimization

The simple mean-variance optimization only requires the expected return vector and expected covariance matrix as inputs. Factors such as each individual's preference become irrelevant. The model is based on a formal quantitative objective that will always give the same solution with the same set of parameters. So the model is not subject to investors' biases due to current or past market events. Also, the formation can be solved efficiently either in closed form or through numerical methods. These all explain its popularity and its contribution to modern portfolio theory (MPT).

However, some of the underlying assumptions of mean-variance portfolio optimization are open to question. For example, the utility function might involve preferences for more than the mean and variance of the portfolio returns and might be a complex function in which a quadratic approximation is not appropriate. Financial asset returns often do not follow normal distribution. Instead, they are often skewed and have fat tails. When the asset return is skewed, it is also arguable whether variance is the correct risk measure since it equally penalizes desirable upside and undersirable downside deviations from the mean. Two alternative measures of risk, semivariance and shortfall risk are sometimes used. The semi-variance of portfolio return r_p

with mean μ_p is defined as $\sigma_{semi}^2 = E[(\mu_p - r_p)^-]^2$, where the expectation is taken with respect to the distribution of r_p, and where $(\mu_p - r_p)^- = \mu_p - r_p$ if $r_p \le \mu_p$ and *0* otherwise. The shortfall risk of a portfolio is defined as $s_\alpha(w) = \mu_p - E[r_p \mid r_p \le q_\alpha(w)]$ where $q_\alpha(w) = \inf\{z \mid P(r_p \le z) \ge \alpha\}$ is the α-quantile of the portfolio. Besides, the one-period nature of static optimization also does not take dynamic factors into account, so some researchers argue for more complicated models based on stochastic processes and dynamic programming.

However, the most serious problem of the mean-variance efficient frontier is probably the method's instability. The mean-variance frontier is very sensitive to the inputs, and these inputs are subject to random errors in estimation of expected return and covariance. Small and statistically insignificant changes in these estimates can lead to a significant change in the composition of the efficient frontier. This may lead us to frequently and mistakenly rebalance our portfolio to stay on this elusive efficient frontier, incurring unnecessary transaction costs. The traditional Markowitz portfolio optimization estimates the expected return and the covariance matrix from historical return time series and treats them as true parameters for portfolio selection. This "certainty equivalence" view has long been criticized because of the impact of parameter uncertainty on optimal portfolio selection (*7-9*). The naïve mean-variance approach often leads to extreme portfolio weights (instead of a diversified portfolio as the method anticipates) and dramatic swings in weights when there is a minor change to the expected returns or the covariance matrix. As a result, the practical application of the mean-variance optimization is seriously hindered by estimation errors. In the following chapters, we will discuss a variety of more sophisticated approaches to address the estimation error problem and to increase the risk-adjusted portfolio performance.

Chapter 2

Shrinkage Models and GARCH Models

2.1 Introduction

As discussed in Chapter 1, the correct estimation of expected returns and the covariance matrix is a crucial step for asset management. Small estimation errors of either input usually lead to a portfolio far from the true optimal efficient frontier. Besides asset allocation, the expected returns and the covariance are also widely used in risk management (e.g. value at risk), derivative pricing, hedging strategies and tests of asset pricing models. Naturally extensive studies have been conducted on their estimation and a variety of methods have been published to tackle the estimation error problem from different angles. Many of the methods address the questions as to how much structure we should place on the estimation. Since equally-weighted sample mean and covariance matrix have the desired property of being unbiased (the expected value is equal to the true covariance matrix) and are easy to compute, they still are the most widely used estimators in asset management. Despite the simple computation involved, this model has high complexity (large number of parameters), so the model suffers from the problem of high variance, which means the estimation errors can be significant and generate erroneous mean-variance efficient frontiers. The problem is further exacerbated if the number of observations is of the same order as the number of assets, which is often the case in financial applications to select industry sectors or individual securities.

Besides the "data-based" approach, which assumes a functional form for the distribution of asset returns and estimates its parameters from the time series of returns, an alternative "model-based" approach imposes a strong structure on the returns, assuming the returns are explained by a number of risk factors. A typical example is the single-factor model based on CAPM, where the expected returns and covariance of assets are believed to be determined by the market beta. This model has low variance in estimation, yet the model may fail to capture the complexity of the relationship between different assets and, as a result, the estimation can be severely biased. More recent research has focused on finding the right tradeoff between the sample covariance matrix and a highly-structured estimator using shrinkage methods, which will be investigated in the first half of this Chapter.

Neither the market model approaches nor the shrinkage methods address the well documented facts that high-frequency financial data tend to substantially deviate from the Gaussian distribution, and the expected returns/covariance matrix is influenced by time dependent information flows. Time series models, especially exponentially weighted moving average models and GARCH models, extract more information from the time series and have the potential to rectify (at least part of) the problem. So in the second half of this Chapter, we will discuss and implement some of the multivariate GARCH models developed over the past decade and evaluate their potential values in asset management.

2.2 Sample Mean and Covariance Matrix

For all our studies, we denote R as a $T \times N$ matrix, where each column vector $R_{\bullet i}$, $i = 1, \cdots, N$ represents a variable and each row $R_{t\bullet}, t = 1, \cdots, T$ represents a cross-sectional observation. For historical return data, each column $R_{\bullet i}$ represents the returns of asset i over different periods, each row $R_{t\bullet}$ represents the returns of different assets at period t. For convenience, we also denote $R_t = R_{t\bullet}^{'}$ as a $N \times 1$ column vector for returns of different assets at period t and r_{ti} (or $r_{t,i}$) as the return of asset i at period t.

The simple sample mean and covariance matrix are the best unbiased estimators using the maximum likelihood method under the assumption of multivariate normality. The covariance matrix can be calculated as

$$S = \frac{1}{T-1}(X - e\bar{R}')'(X - e\bar{R}') = \frac{1}{T-1}R'\left(I - \frac{1}{T}ee'\right)R \tag{2.1}$$

where $\bar{R}$ is a $N \times 1$ sample mean vector with $\bar{R}_i = \frac{1}{T}\sum_{t=1}^{T} r_{ti}$; e is a $T \times 1$ vector of ones; I is a $T \times T$ identity matrix;

Besides its intuitive interpretation, the sample mean and the sample covariance matrix require trivial computation. However the number of the parameters estimated is $N(N+3)/2$. When the number of assets is large relative to the number of historical return observations available, the sample covariance is often estimated with a lot of error. One natural solution to address this problem is to use a long historical period with a large number of samples. But distant past performance may not represent the future population distribution since the underlying (economic, market, regulatory, and psychological) fundamentals could well have changed over time. Alternative solutions are factor models that impose structure on the estimation.

2.3 Factor Models

Factor models assume the asset returns are generated by specific factors. These factors can either be observable economical factors or principal components extracted from the data. The fewer the number of factors we use, the stronger the structure we impose on the estimation.

2.3.1 Single-factor Model

The capital asset pricing model (CAPM) , which was developed by Sharpe in 1963 (*5, 10*), explains the asset returns using a single factor: the market return. The CAPM model for the an individual asset is expressed as

$$r_{ti} = \alpha_i + \beta_i r_{tm} + \varepsilon_{ti} \tag{2.2}$$

where r_{ti} is the return of asset i for period t; α_i is the intercept of regression for asset i; β_i is the market beta for asset i; r_{tm} is the market return for period t; ε_{ti} is the residual.

If we assume that the residuals ε_{ti} are uncorrelated with each other and with market return r_{tm}, the covariance for each pair of assets only correlate with each other through the market return and the covariance matrix of all assets can be expressed as

$$F = \sigma_m^2 \beta\beta' + E \tag{2.3}$$

where F is the covariance matrix for x_is $i = 1, 2, \cdots, N$; σ_m^2 is the variance of market returns; β is the market beta; E is the residual covariance matrix, $E[\varepsilon\varepsilon'] = Diag\left(\sigma_{1,\varepsilon}^2, \cdots, \sigma_{n,\varepsilon}^2\right)$;

In practice, none of the parameters σ_m^2, β, or E is observable. They are usually estimated from simple linear regression and the corresponding covariance matrix can be estimated as

$F = \hat{\sigma}_m^2 \hat{\beta}\hat{\beta}' + \hat{E}$. The market return can be estimated by the market-value weighted return (as originally proposed in the CAPM model) using $r_{tm} = \sum_{i=1}^{N} M_{ti} r_{ti} / \sum_{i=1}^{N} M_{ti}$, where M_{ti} is the market value of asset *i* at the beginning of period *t*. The market model only requires estimation of $2N+1$ parameters ($\hat{\beta}_i$ s, $\hat{\sigma}_{i,\varepsilon}^2$ s and $\hat{\sigma}_m^2$), which is a large reduction compared with $N(N+3)/2$ parameters for the sample mean and covariance matrix. As a result, the method has less estimation variance. However it replaces estimation variance with large specification error (bias) since market returns cannot explain all of the variance of asset returns.

2.3.2 Multi-factor (Barra) model

To reduce the specification error, multi-factor models are often used to explain asset returns and the formulation of the an individual asset's return is extended to

$$r_{ti} = \alpha_i + \beta_{i1}\Lambda_{t1} + \cdots + \beta_{iK}\Lambda_{tK} + \varepsilon_{ti} \tag{2.4}$$

where r_{ti} is the return of asset *i* for period *t*; α_i is the intercept of regression for asset *i*; β_{ik} is the coefficient of the *k*th factor for asset *i*, $k = 1, \cdots, K$ Λ_{tk} is the 'return' of the *k*th factor for period *t*, $k = 1, \cdots, K$; ε_{ti} is the residual.

The corresponding estimation for covariance matrix becomes

$$F = B\Omega B' + E \tag{2.5}$$

where B is a $N \times K$ matrix, the coefficient matrix for all N asset classes; Ω is a $K \times K$ matrix, the covariance matrix for the K factors; E is the residual covariance matrix, $E[\varepsilon\varepsilon'] = Diag(\sigma_1^2, \cdots, \sigma_n^2)$.

Unlike the sample covariance matrix, the factor models using K factors only need to estimate $NK + K(K+1)/2 + N$ (for B, Ω and E respectively) parameters, which is much smaller than $N(N+3)/2$ since the number of explanatory variables is usually much smaller than the number of assets ($K << N$).

The factors used in these models are often variables that have proven their explanatory power in equity research. Financial research companies such as BARRA and APT offer a wide variety of factors for risk indexes as well as country and industry dummy variables, so linear factor models based on BARRA or APT factors are often used to construct the covariance matrix. Nevertheless the implementation of these models depends on costly external proprietary data and the validity of those factors is not open for independent verification. The success of estimation relies on the correct choices of the factors and the validity of the estimation of both the factor and factor coefficients for different assets. In practice, both the choices of factors and estimation of coefficients have been subjected to debate. To avoid using these external economic, financial and industrial factors, principal components derived from the sample covariance matrix can be used as alternative factors.

2.3.3. Principal component analysis (PCA)

Principal component analysis assumes that the covariance matrix can be explained by a few linear combinations of the original variables. Apply the singular value decomposition (SVD) to the covariance matrix:

$$\Sigma = UDV^T \Leftrightarrow \Sigma V = VD \tag{2.6}$$

where U is a $T \times N$ matrix with orthogonal columns, $U^T U = I$; D is a diagonal $N \times N$ matrix;

V is a $N \times N$ matrix with orthogonal columns, $V^T V = I$; The diagonal element of D and columns of V are eigenvalue-eigenvector pairs (V_1, λ_1), $\cdots$, (V_N, λ_N) where $\lambda_1 \geq \cdots \geq \lambda_N \geq 0$

Instead of an exact representation using all eigenvalue-eigenvector pairs, we can estimate Σ using the first K pairs, and the approximation can be expressed as

$$\Sigma \approx V(K) \times D(K) \times V(K)' + \hat{E} \tag{2.7}$$

where $V(K)$ is a $N \times K$ which represents the first K columns of V; $D(K)$ is a $K \times K$ matrix which represents the top-left sub-matrix of D; $\hat{E}$ is the residual covariance matrix, $E[\varepsilon\varepsilon'] = Diag\left(\sigma_{1,\varepsilon}^2, \cdots, \sigma_{n,\varepsilon}^2\right)$

A model using K principal components only needs to estimate $NK + K + N$ (for $V(K)$, $D(K)$, and E respectively) parameters, which is similar to multi-factor models. The difference is that instead of using external factors, the factors are extracted as principal components, which are linear combinations of the return variables. So there is no reliance on external sources for factors. The drawback is that principal components may not have clear interpretations, although it is widely believed that the first principal component represents a kind of market factor and others often mirror industry specific effects.

2.4. Shrinkage methods

To further pursue a tradeoff between estimation error and specification error, we can also combine the simple sample covariance matrix with factor models using shrinkage methods. Shrinkage, one of the Bayesian statistical approaches, assumes a prior which should reduce the dependency on purely estimated parameters by representing some form of structure. One of the popular shrinkage methods used in statistics is ridge regression, which has a prior that all regression coefficients $\beta = 0$. The covariance matrix estimated from a factor model can be used as prior and the goal is to calculate a weighted average of highly structured single-index/K-factor-index model covariance matrix (F) and the sample covariance matrix (S):

$$\hat{\Sigma} = \alpha F + (1-\alpha)S \qquad (2.8)$$

where α is the shrinkage constant with constraint $0 < \alpha < 1$.

To generate the final covariance matrix, we need to choose an optimal shrinkage constant. The optimal shrinkage constant is often chosen as the scalar α $(0 < \alpha < 1)$ that minimizes the loss function $\sum_{t=1}^{T}(R_t - \mu)'\hat{\Sigma}^{-1}(R_t - \mu)$, which is the sum of Mahalanobis distance of all data points. This optimization problem has only one variable λ, which can be easily solved using a variety of optimization algorithms. The problem is that this loss function depends on the inverse of the covariance matrix, which can become singular when $N \geq T$. Even when N < T, the matrix may be ill-conditioned, and as a result the inverse introduces much calculation error.

Ledoit and Wolf (*11*) proposed another intuitive loss function: a quadratic measure of distance between the true and the estimated covariance matrices based on the Frobenius norm instead of Mahalanobis distance. The Frobenius norm of the $N \times N$ symmetric matrix Z with eigenvalues λ_i $(i = 1, \cdots, N)$ is defined by:

$$\|Z\|^2 = Trace\left(Z^2\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} z_{ij}^2 = \sum_{i=1}^{N} \lambda_i^2 . \tag{2.9}$$

So we choose an 'optimal' α to minimize the following loss function based on Frobenius norm:

$$L(\alpha) = \|\alpha F + (1-\alpha)S - \Sigma\|^2 , \tag{2.10}$$

where Σ is the true covariance matrix. This approach does not depend on the inverse of the estimated covariance matrix and avoids the singularity problem even when $N \geq T$. Yet the choice of α requires the true covariance matrix, which is unobservable from the data. Instead the authors came up with an asymptotic solution to this shrinkage problem:

If we take the expectation Frobenius norm loss function $L(a)$

$$\begin{aligned} R(a) = E\left[L(a)\right] &= \sum_{i=1}^{N}\sum_{i=1}^{N} E(af_{ij} + (1-a)s_{ij} - \sigma_{ij})^2 \\ &= \sum_{i=1}^{N}\sum_{i=1}^{N} \mathrm{var}(af_{ij} + (1-a)s_{ij} - \sigma_{ij}) + \left[E\left(af_{ij} + (1-a)s_{ij} - \sigma_{ij}\right)\right]^2 \\ &= \sum_{i=1}^{N}\sum_{i=1}^{N} \mathrm{var}(af_{ij} + (1-a)s_{ij} - \sigma_{ij}) + \left[E\left(a(f_{ij} - \sigma_{ij}) + (1-a)(s_{ij} - \sigma_{ij})\right)\right]^2 \\ &= \sum_{i=1}^{N}\sum_{i=1}^{N} a^2 \,\mathrm{var}(f_{ij}) + (1-a)^2 \,\mathrm{var}(s_{ij}) + 2a(1-a)\,\mathrm{cov}(f_{ij}, s_{ij}) + a^2(\phi_{ij} - \sigma_{ij})^2 \end{aligned} \tag{2.11}$$

where f_{ij} is the estimated covariance of *i*th and *j*th asset using factor model; s_{ij} is the estimated covariance of *i*th and *j*th asset using sample covariance matrix; σ_{ij} is the true covariance of *i*th and *j*th asset; ϕ_{ij} is the expected value of f_{ij} .

To minimize $R(a)$ with respect to a, we set $R'(a)=0$

$$\left.\begin{aligned} R'(a) &= 2\sum_{i=1}^{N}\sum_{i=1}^{N} a\operatorname{var}(f_{ij}) - (1-a)\operatorname{var}(s_{ij}) + (1-2a)\operatorname{cov}(f_{ij},s_{ij}) + a(\phi_{ij}-\sigma_{ij})^2 \\ R''(a) &= 2\sum_{i=1}^{N}\sum_{i=1}^{N} \operatorname{var}(f_{ij}-s_{ij}) + (\phi_{ij}-\sigma_{ij})^2 \geq 0 \end{aligned}\right\} \Rightarrow$$

$$a^* = \frac{\sum_{i=1}^{N}\sum_{i=1}^{N} \operatorname{var}(s_{ij}) - \operatorname{cov}(f_{ij},s_{ij})}{\sum_{i=1}^{N}\sum_{i=1}^{N} \operatorname{var}(f_{ij}-s_{ij}) + (\phi_{ij}-\sigma_{ij})^2} \tag{2.12}$$

Asymptotically, the optimal shrinkage constant α is $a^* = \frac{1}{T}\frac{\pi-\rho}{\gamma} + O\left(\frac{1}{T^2}\right)$, where π is the sum of asymptotic variances of the entries of the sample covariance matrix scaled by $\sqrt{T}$ $\pi = \sum_{i=1}^{N}\sum_{i=1}^{N} Asy\operatorname{var}(\sqrt{T}S_{ij})$; ρ is the sum of asymptotic covariances of the entries of the factor-model covariance matrix with the entries of sample covariance matrix scaled by $\sqrt{T}$: $\rho = \sum_{i=1}^{N}\sum_{i=1}^{N} \rho_{ij} = \sum_{i=1}^{N}\sum_{i=1}^{N} Asy\operatorname{cov}\left(\sqrt{T}F_{ij},\sqrt{T}S_{ij}\right)$; γ is the misspecification of the factor-model: $\gamma = \sum_{i=1}^{N}\sum_{i=1}^{N}\left(\phi_{ij}-\sigma_{ij}\right)^2$.

So it is clear that the weight (α) places on the factor-model increases with the variances of the estimated sample covariance matrix (through π) and decreases with misspecification of the factor model (through γ), which is consistent with the Bayesian principle.

In practice, none of the asymptotic variances π, asymptotic covariance ρ, and misspecification γ are observable. Nevertheless consistent estimators can be calculated as:

$$\hat{\pi}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\left((r_{ti}-\overline{r}_i)(r_{tj}-\overline{r}_j)-s_{ij}\right)^2$$

$$\hat{\rho}_{ijt} = \frac{s_{jm}s_{mm}r_{ti}+s_{im}s_{mm}r_{tj}-s_{im}s_{jm}r_{tm}}{s_{mm}^2}r_{tm}r_{ti}r_{tj} - f_{ij}s_{ij} \tag{2.13}$$

$$\hat{\rho}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\hat{\rho}_{ijt} \text{ and } \hat{\gamma}_{ij} = (f_{ij}-s_{ij})^2$$

where $\overline{r}_i$ is the mean return of asset i; f_{ij} is the estimated covariance of ith and jth asset using factor model; s_{ij} is the estimated covariance of ith and jth asset (including market as expressed in m) using the sample covariance matrix;

Ledoit *et al.* applied the asymptotic solution to data from Center for Research in Security Prices (CRSP) monthly data, which includes 909 assets and 23 years (1972 – 1995). 10-year historical data (120 months) was used to estimate the sample covariance matrix and CAPM-based covariance matrix. The shrinkage estimator was then estimated using the asymptotic approximation. To avoid the complication of expected return estimation, the model was applied to global minimum variance portfolio and the portfolio was updated monthly. The results showed that the shrinkage estimator indeed gave lower annualized standard deviation (9.55% v.s 14.27% for sample covariance and 12.00% for CAPM model), however no return information was given for these portfolios.

Interestingly, none of these papers discussed the influence of factor models on expected return estimation. However, if simple LSE regression is used to estimate factor coefficients (as all these papers did), the corresponding estimation of expected return is the same as the sample mean. The reason is that LSE regression yields an unbiased estimator, which indicates that on average the mean of the predicted response variable is equal to the mean of observed response variable. The shrinkage method does not alter this characteristic, so all these methods yield the same expected return estimation as the simple sample mean.

The models discussed so far assign a weight of 1 to every data point that is included and 0 otherwise. Nevertheless, evidence showed that more recent return data often have larger influence on future returns. The unweighted approaches fail to take the time series properties (e.g. heteroskedasticity and volatility clustering) of the return data into consideration. As a result, these models may not fully utilize all the information included in the data.

2.5. Exponentially Weighted Moving Average (EWMA) Model

One simple and popular method to take into account the heteroskedasticity of financial returns is the exponentially weighted moving average (EWMA) model. It is generally expressed as the following equation:

$$\sigma_{ij,t+1} = (1-\lambda)\sum_{s=0}^{t-1} \lambda^s r_{t-s,i} r_{t-s,j} + \lambda^t \sigma_{ij,0} \quad \Leftrightarrow \quad \sigma_{ij,t+1} = \lambda \sigma_{ij,t} + (1-\lambda) r_{t,i} r_{t,j} \tag{2.14}$$

where λ is the decay factor, $0 < \lambda < 1$; $\sigma_{ij,t}$ and $\sigma_{ij,t+1}$ are the estimated covariances between asset i and j at the beginning of t and $t+1$ respectively; $r_{t,i}$ and $r_{t,j}$ are the returns of asset i and j for period t.

More precisely, the equation is $\sigma_{ij,t+1} = \lambda \sigma_{ij,t} + (1-\lambda)(r_{t,i} - \overline{r}_i)(x_{t,j} - \overline{r}_j)$. Since EWMA is often applied to high-frequency daily data, $\overline{r}_i$ and $\overline{r}_j$ are expected to be relatively small compared with the standard deviation and are omitted in the equation. Such omission results in a bias of approximately $\overline{r}_i \overline{r}_j$, which is usually insignificant for daily returns.

The choice of decay rate has a large influence on the estimated covariance matrix. A lower value of λ gives old data little weight and produces estimates of daily volatility that responds more rapidly to the new daily returns. The RiskMetrics database developed by J.P. Morgan use λ = 0.94 to update daily volatility estimates for US market. In general we can define an optimal decay factor to minimize the following loss function: $\min_{\lambda} \sum_{t=1}^{T} \left\| \Omega_t(\lambda) - V_t \right\|$ where T is number of periods; $\Omega_t(\lambda)$ is the covariance matrix calculated using EWMA model for period t; V_t is the historical covariance for period t, with (i, j)th element $r_{t,i}r_{t,j}$; $\|\bullet\|$ is the Frobenius norm. (This is an extension of the RiskMetrics approach. No exact reference for the estimation has been found because the information is proprietary, so this generalization is an educated guess.) Alternatively, we can optimize λ_{ij} for each pair of assets, yet it means estimating an extra N^2 parameters (which raises the possibility of over fitting) and there is no strong economic reasoning to do so. Another pitfall is that the calculated covariance matrix may not be positive semidefinite.

A different optimization method estimates the optimal λ using the maximum likelihood method by assuming a conditional multivariate normal distribution of the returns. Nevertheless it is well established that financial returns, especially high-frequency returns, do not follow a normal distribution. Instead, many asset returns have fat tails compared with the normal distribution. To address the fat-tail problem in financial returns, a mixture of normal distributions was used by Goldman Sachs to make fat-tailed returns more likely. The method assumes that most of the time return vectors are drawn from a low volatility state; but with a small probability that returns are drawn from a high volatility state. Furthermore, the method assumes that the variance of the high volatility state is a constant multiple of the variance of the low volatility state. For example, the probability density function of a single asset at period t can be expressed as:

$$f(r_t,\sigma_t^2)=p\times\frac{1}{\sqrt{2\pi}\sigma_{ts}}e^{-r_t^2/\sigma_{ts}^2}+(1-p)\times\frac{1}{\sqrt{2\pi}\sigma_{th}}e^{-r_t^2/\left(\sigma_{th}^2\right)} \tag{2.15}$$

$$s.t.\quad \sigma_t^2=p\times\sigma_{ts}^2+(1-p)\times\sigma_{th}^2,\ \sigma_{th}^2=k\sigma_{ts}^2$$

where r_t is the return of the asset at time t; σ_{ts}^2 is the variance of the low volatility state at time t ;σ_h^2 is the variance of the high volatility state at time t; σ_t^2 is the average volatility at time t; k is a constant, $k>1$.

Now the likelihood function depends on both p (the probability of the low volatility state) and k. For each λ, we can find values p and k that maximize the likelihood function. The optimal decay rate λ_{opt} that maximizes the log-likelihood ratio $L(r_t,t=1,\cdots,T)=\prod_{t=1}^{T}f(r_t,\sigma_t^2)$, where $\sigma_{t+1}^2=\lambda\sigma_t^2+(1-\lambda)r_t^2$, is identified by repeating this process for many values of λ.

2.6. GARCH Model

As we have discussed in the previous section, probability distributions for asset returns often exhibit fatter tails than the normal distribution and are referred to as leptokurtic. In addition, return time series usually exhibit characteristics such as volatility clustering (in which large changes tend to follow large changes, and small changes tend to follow small changes) and leverage effects (negative correlation between asset returns and volatility). Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models were developed by Bollerslev in 1986 (*12*) and are arguably the most popular methods to estimate conditional variances.

GARCH models not only address the changes of variance/covariance over time, but also account for leptokurtosis, volatility clustering and leverage effects.

We begin with a univariate GARCH(1, 1) model for conditional variance estimation since it is a well established method shown to yield good results in many studies. The conditional variance of asset i that follows a univariate GARCH(1, 1) model is given as a function of the long-term unconditional variance, the squared return at time t and the estimated variance at time t:

$$\sigma_{ii,t+1} = (1-\alpha_{ii}-\beta_{ii})V_{ii,L} + \alpha_{ii}r_{t,i}^2 + \beta_{ii}\sigma_{ii,t} = \gamma_{ii}V_{ii,L} + \alpha_{ii}r_{t,i}^2 + \beta_{ii}\sigma_{ii,t} = \omega_{ii} + \alpha_{ii}r_{t,i}^2 + \beta_{ii}\sigma_{ii,t}$$

$$s.t. \quad \alpha_{ii} \geq 0, \beta_{ii} \geq 0, \gamma_{ii} > 0 \qquad (2.16)$$

where $V_{ii,L}$ is the long-term unconditional variance; $r_{t,i}^2$ is the squared return at time t; $\sigma_{ii,t}$ is the estimated variance at time t; α_{ii}, β_{ii} and γ_{ii} are the weight constants for $r_{t,i}^2$, $\sigma_{ii,t}$ and $V_{ii,L}$ respectively with $\gamma_{ii} = 1-\alpha_{ii}-\beta_{ii}$; $\omega_{ii} = \gamma_{ii}V_{ii,L} > 0$.

The implicit constraint is $\alpha_{ii}+\beta_{ii} < 1$, which is used to guarantee that the unconditional variance is stationary. To estimate the parameters α_{ii}, β_{ii}, and ω_{ii}, we solve the maximum likelihood problem by assuming conditional normality:

$$\max_{\omega_{ii},\alpha_{ii}+\beta_{ii}\leq 1-\varepsilon} \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_{ii,t}}} \exp\left[-\tfrac{1}{2}r_{t,i}^2/\sigma_{ii,t}\right] \qquad (2.17)$$

$$s.t. \quad \sigma_{ii,t+1} = \omega_{ii} + \alpha_{ii}r_{t,i}^2 + \beta_{ii}\sigma_{ii,t}, \alpha_{ii} \geq 0, \beta_{ii} \geq 0, \omega_{ii} > 0$$

Multivariate GARCH models (*13-15*) have also been extensively investigated in econometric literature and are used by some sophisticated practitioners. The most general multivariate GARCH model can be defined as:

$$\begin{cases} E\left[r_{t+1,i} \mid F_t\right] = 0 \\ E\left[\sigma_{ij,t+1} \mid F_t\right] = \gamma_{ij} V_{ij,L} + \alpha_{ij} r_{t,i} r_{t,j} + \alpha_{ij} \sigma_{ij,t} = \omega_{ij} + \alpha_{ij} r_{t,i} r_{t,j} + \beta_{ij} \sigma_{ij,t} \end{cases} \tag{2.18}$$

$$s.t. \quad \alpha_{ij} + \beta_{ij} + \gamma_{ij} = 1, \alpha_{ij} \geq 0, \beta_{ij} \geq 0, \gamma_{ij} > 0 (\omega_{ij} > 0)$$

where F_t is the information available at *t*; $V_{ij,L}$ is the long-term expected covariance between asset *i* and *j*. $i, j = 1, \cdots, N$; $r_{t,i}$ and $r_{t,j}$ are the returns of asset *i* and *j* for period *t*; $\sigma_{ij,t}$ is the estimated covariance between asset *i* and *j* at time *t*.

The estimated parameters for all *i* and *j* can be expressed using three matrices:

$$A = \left[\alpha_{ij}\right], B = \left[\beta_{ij}\right], C = \left[\omega_{ij}\right] \quad i, j = 1, \cdots, N$$

and the conditional covariance for *t+1* can be expressed as:

$$\Sigma_{t+1} = C + A \cdot *(R_t R_t') + B \cdot * \Sigma_t \tag{2.19}$$

where $A \cdot *(R_t R_t')$ is the element-by-element product of matrix A and $R_t R_t'$; $B \cdot * \Sigma_t$ is the element-by-element product of matrix B and Σ_t; $R_t R_t'$ is the matrix of cross-products of returns observed at time *t*; Σ_{t+1} and Σ_t are the conditional covariance matrix at time $t+1$ and t respectively.

In the research literature, the Vech operator is often used to transform a symmetric matrix into a stack vector:

$$\Sigma_t = \begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} & \cdots & \sigma_{1N,t} \\ \sigma_{12,t} & \sigma_{22,t} & \cdots & \sigma_{2N,t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1N,t} & \sigma_{2N,t} & \cdots & \sigma_{NN,t} \end{bmatrix} \Rightarrow vech(\Sigma_t) = \begin{bmatrix} \sigma_{11,t} \\ \vdots \\ \sigma_{1N,t} \\ \sigma_{22,t} \\ \sigma_{23,t} \\ \vdots \end{bmatrix} \tag{2.20}$$

So the conditional covariance for t+1 can be expressed as:

$$vech(\Sigma_{t+1}) = \begin{bmatrix} \sigma_{11,t+1} \\ \vdots \\ \sigma_{1N,t+1} \\ \sigma_{22,t+1} \\ \sigma_{23,t+1} \\ \vdots \end{bmatrix} = \begin{bmatrix} \omega_{11} \\ \vdots \\ \omega_{1N} \\ \omega_{22} \\ \omega_{23} \\ \vdots \end{bmatrix} + \begin{bmatrix} \alpha_{11} & 0 & & \cdots & & & 0 \\ 0 & \ddots & 0 & & & & \\ 0 & \cdots & \alpha_{1N} & 0 & \cdots & & \\ \vdots & 0 & 0 & \alpha_{22} & 0 & \cdots & \\ 0 & \cdots & & 0 & \alpha_{23} & \cdots & \\ 0 & & \cdots & \vdots & \vdots & \ddots & \end{bmatrix} \times \begin{bmatrix} r_{1,t}^2 \\ \vdots \\ r_{1,t}r_{N,t} \\ r_{2,t}^2 \\ r_{2,t}r_{3,t} \\ \vdots \end{bmatrix} + \begin{bmatrix} \beta_{11} & 0 & & \cdots & & & 0 \\ 0 & \ddots & 0 & & & & \\ 0 & \cdots & \beta_{1N} & 0 & \cdots & & \\ \vdots & 0 & 0 & \beta_{22} & 0 & \cdots & \\ 0 & \cdots & & 0 & \beta_{23} & \cdots & \\ 0 & & \cdots & \vdots & \vdots & \ddots & \end{bmatrix} \begin{bmatrix} \sigma_{11,t} \\ \vdots \\ \sigma_{1N,t} \\ \sigma_{22,t} \\ \sigma_{23,t} \\ \vdots \end{bmatrix}$$

This model has $3N(N+1)/2$ parameters and the parameters are selected by maximizing the likelihood ratio. This equation is widely known as Diagonal-Vech model. The estimation of these $3N(N+1)/2$ parameters requires the optimization of all the parameters α_{ij}, β_{ij}, and ω_{ij} simultaneously using the conditional maximum likelihood estimation method. The log likelihood function is usually very flat near the optimum, which makes gradient methods converge slowly. As a result, the computational time often increases exponentially with the number of assets. Even for a moderate number of assets, the problem may become intractable.

Another pitfall is that the parameters are supposed to have interactions with each other. For example, if variable a has high correlation with variable b and variable c, then b and c are supposed to have high correlation as well. A naively constructed covariance matrix from the Diagonal-Vech model may not be positive semi-definite. A few models have been developed to guarantee positive semi-definiteness of the matrix. These models mainly differ on three aspects:

1. The number of parameters: some of the models try to keep the number of free parameters low by expressing some of the parameters as functions of other parameters. Certainly the validity of these models often depends on the validity of the assumptions. So it is a tradeoff between dimensionality and validity of assumption.
2. The restrictions on the parameter space: some models place explicit restrictions on the parameter space to guarantee the positive semi-definite nature of the covariance matrix. The risk of doing that is to impose restrictions that may be strongly violated by the data. Others start without much restriction and convert the covariance matrix to guarantee the positive semi-definite nature through numerical methods. Yet such approaches often lack strong theoretical explanations. So it is a tradeoff between validity and interpretability of the restrictions.
3. Correlation vs. covariance: Some of the models estimate the correlation coefficient matrix first and derive the covariance from the correlation coefficient matrix and the conditional variance of the each asset, which can be estimated using univariate GARCH.

We will discuss some of the popular methods developed by economists and apply them to our asset allocation problems. In practice, the choice of a particular model often depends on empirical results and the experience of each individual user.

2.6.1. Constant Conditional Correlation GARCH

Bollerslev (*16*) suggested an intuitive approach to guarantee that the covariance will be positive semi-definite by using constant conditional correlation assumption. The conditional variance of each asset $\sigma_{ii,t}$ is modeled as a univariate GARCH(1, 1)model and the corresponding parameters α_{ii}, β_{ii}, and ω_{ii} are estimated. The conditional covariance between each variable pair R_i and R_j

is calculated as $\sigma_{ij,t} = \rho_{ij}\sqrt{\sigma_{ii,t}\sigma_{jj,t}}$. For a consistent conditional correlation coefficient matrix

$$\Gamma = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1N} \\ \rho_{12} & 1 & \cdots & \rho_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1N} & \rho_{2N} & \cdots & 1 \end{bmatrix} \text{ and stochastic diagonal matrix } D_t = \begin{bmatrix} \sqrt{\sigma_{11,t}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22,t}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{NN,t}} \end{bmatrix},$$

the Conditional covariance matrix can be expressed as $\Sigma_t = D_t \Gamma D_t$. It is worth noting that the conditional coefficient matrix Γ instead of the unconditional one is supposed to be used. In Bollerslev's paper, the estimation of Γ was not apparent. One sensible approach is to estimate $\sigma_{ii,t}$, $\forall i = 1,\cdots,N$ *and* $\forall t = 1,\cdots,T$, standardize the returns by $\tilde{r}_{t,i} = r_{t,i} / \sqrt{\sigma_{ii,t}}$ and then use the standardized returns' pair-wise sample correlation coefficient to estimate the ρ_{ij} s.

The method reduces the number of parameters to $N(N+5)/2$ and only N univariate GARCH models need to be performed in order to yield α_{ii}, β_{ii}, and ω_{ii}. The $N(N+1)/2$ parameters in the correlation matrix can be computationally trivially estimated as the sample correlation matrix. So the model is the simplest among all multivariate GARCH model and the computation time only increases linearly with the number of assets. The disadvantage of the model is clearly the constant correlation assumption since the conditional correlations may vary over time.

2.6.2 Dynamic Conditional Correlation GARCH

To incorporate the changes of conditional correlations over time, Tse and Engle proposed a generalization of the conditional correlation GARCH by making the conditional correlation matrix Γ_t time dependent, or $\Sigma_t = D_t \Gamma_t D_t$ (*14, 17, 18*).

The standardized disturbance is $\varepsilon_t = D_t^{-1} R_t$ and the conditional correlation matrix can be expressed as $\Gamma_t = D_t^{-1} \Sigma D_t^{-1} = E_{t-1}(\varepsilon_t \varepsilon_t')$. The simplest approach to estimate the correlation matrix is to use an exponentially weighted average model, which is similar to what we have discussed for covariance matrix. The process can be expressed as:

$$\Gamma_{t+1} = \lambda \Gamma_t + (1-\lambda)\varepsilon_t \varepsilon_t' \tag{2.21}$$

Alternatively, we can apply the Scalar GARCH(1, 1) model to the conditional correlation matrix:

$$\Gamma_{t+1} = (1-\alpha-\beta)\bar{\Gamma} + \alpha \varepsilon_t \varepsilon_t' + \beta \Gamma_t \tag{2.22}$$

The diagonal elements of the conditional correlation are 1's, so the total number of parameters is $(N+1)(N+4)/2$. The implementation of Dynamic Conditional Correlation GARCH can be formulated as a two step process. Consider the log likelihood function:

$$\begin{aligned} L &= -\tfrac{TN}{2}\log(2\pi) - \tfrac{1}{2}\sum_{t=1}^{T}\left(\log|\Sigma_t| + R_t'\Sigma^{-1}R_t\right) \\ &= -\tfrac{TN}{2}\log(2\pi) - \tfrac{1}{2}\sum_{t=1}^{T}\left(\log|D_t\Gamma_t D_t| + R_t' D_t^{-1}\Gamma_t^{-1}D_t^{-1}R_t\right) \\ &= -\tfrac{TN}{2}\log(2\pi) - \tfrac{T}{2}\log\left(|D_t|\right) - \tfrac{1}{2}\sum_{t=1}^{T}\log|\Gamma_t| - \tfrac{1}{2}\sum_{t=1}^{T}\varepsilon_t'\Gamma_t\varepsilon_t \end{aligned} \tag{2.23}$$

In this equation, there are only two components that can vary. The first part contains only D_t, the second part only Γ_t (conditioned on D_t). So we can find α_{ii}, β_{ii}, and ω_{ii} using N univariate GARCH models, which will give us D_t. In the second step, the parameters $\bar{\Gamma}, \alpha$ and β are estimated to construct the conditional correlation matrix. Since the parameters of the variance estimation and correlation matrix estimation are not simultaneous, the procedure is statistically inefficient, yet it is consistent. Besides its simplicity, this approach implicitly guarantees that the conditional correlation matrix is positive semi-definite. Because $\bar{\Gamma}$, $\varepsilon_t\varepsilon_t'$ and Γ_t are all positive semi-definite, Γ_{t+1} is also positive semi-definite. The drawback of the model is that α and β are scalars, so the model imposes the restriction that all the conditional correlations obey the same dynamics, which may not be true in reality.

2.6.3. BEKK GARCH

To guarantee the positive semidefiniteness of the covariance matrix, BEKK GARCH(1, 1) diagonal model (based on work by Baba, Engle, Kraft and Kroner) was developed. The general model is defined as:

$$\Sigma_{t+1} = G'G + E'(R_tR_t')E + F'\Sigma_tF \tag{2.27}$$

where G is a triangular matrix; E and F are diagonal matrices, $e_{ii}^2 + f_{ii}^2 < 1 \quad \forall i = 1,\cdots,N$. The restriction $e_{ii}^2 + f_{ii}^2 < 1 \quad \forall i = 1,\cdots,N$ guarantees that the conditional covariance matrix is stationary. The triangular matrix G and diagonal matrices E and F guarantees that the final covariance matrix is positive semidefinite. The method estimates a total of $N(N+5)/2$ parameters. The BEKK GARCH(1, 1) model has the implicit assumption that $a_{ij} = \sqrt{a_{ii}a_{jj}}$ and $b_{ij} = \sqrt{b_{ii}b_{jj}}$, which may not hold, so it is the constraint of the model.

2.6.4 Scalar GARCH

The Scalar GARCH method sets all $\alpha_{ij} = \alpha \quad \forall i, j = 1,\cdots,N$ and $\beta_{i,j} = \beta \quad \forall ij = 1,\cdots,N$, so the BEKK problem becomes:

$$vech(\Sigma_{t+1}) = \begin{bmatrix} \sigma_{11,t+1} \\ \vdots \\ \sigma_{1N,t+1} \\ \sigma_{22,t+1} \\ \sigma_{23,t+1} \\ \vdots \end{bmatrix} = \begin{bmatrix} \omega_{11} \\ \vdots \\ \omega_{1N} \\ \omega_{22} \\ \omega_{23} \\ \vdots \end{bmatrix} + \begin{bmatrix} \alpha & 0 & & \cdots & & 0 \\ 0 & \ddots & 0 & & & \\ 0 & \cdots & \alpha & 0 & \cdots & \\ \vdots & 0 & 0 & \alpha & 0 & \cdots \\ 0 & \cdots & & 0 & \alpha & \cdots \\ 0 & & \cdots & \vdots & \vdots & \ddots \end{bmatrix} \times \begin{bmatrix} r_{1,t}^2 \\ \vdots \\ r_{1,t}r_{N,t} \\ r_{2,t}^2 \\ r_{2,t}r_{3,t} \\ \vdots \end{bmatrix} + \begin{bmatrix} \beta & 0 & & \cdots & & 0 \\ 0 & \ddots & 0 & & & \\ 0 & \cdots & \beta & 0 & \cdots & \\ \vdots & 0 & 0 & \beta & 0 & \cdots \\ 0 & \cdots & & 0 & \beta & \cdots \\ 0 & & \cdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \sigma_{11,t} \\ \vdots \\ \sigma_{1N,t} \\ \sigma_{22,t} \\ \sigma_{23,t} \\ \vdots \end{bmatrix} \quad (2.26)$$

The method also reduces the number of parameters to $N(N+1)/2+2$ and the parameters again can be estimated using maximum likelihood method.

Engle and Mezrich also provided a useful tool to reduce the number of parameters to merely 2. Consider the basic equation estimation $\sigma_{ij,t+1} = (1-\alpha_{ij}-\beta_{ij})V_{ij,L} + \alpha_{ij}r_{t,i}r_{t,j} + \beta_{ij}\sigma_{ij,t}$, if $V_{ij,L}$ is known, we can set $\omega_{ij} = (1-\alpha_{ij}-\beta_{ij})V_{ij,L}$. By letting $V_{ij,L}$ be the long-term pair-wise unconditional variance of the return vectors, we no longer need to estimate ω_{ij} in the optimization and the number of the total parameters is reduced to 2. This simplification essentially cuts the link between the number of variables and the number of parameters in the optimization, so it can a suitable choice if the number of underlying assets is large.

Both the EWMA and GARCH models are designed to remove autocorrelation of the covariance. If the models work well, $r_{t,i}r_{t,j}/\sigma_{ij,t}$ should show little autocorrelation, which can serve as a test for the effectiveness of these two methods. The major difference is that EWMA models do not include the long-term unconditional variance term for future variance prediction. As a result, GARCH(1,1) models are mean reverting: the tendency for volatility after a period of being unusually high or low to move towards a long run average level, while EWMA models are not. Since real stock return variances tend to be mean reverting, GARCH models could be theoretically superior. However, since the optimal parameters α, β, and γ are estimated and updated using an iterative approach, GARCH models often take significant amounts of computation.

2.7. Application to Historical Data

The application uses daily return data on 51 MSCI US industry sector indexes (see appendix 1 for details), from 01/03/1995 to 02/07/2005, which amounts to 2600 days of data. Combining all these industries together, they form a general index for US equity market broader than S&P 500. So the weighted average returns of these industry sectors are often used as a proxy of market index. In this section, we study the performance of the following estimators:

Table 2.1. Application of shrinkage and GARCH methods

Method	Expected Return Estimation	Covariance matrix Estimation
V	Sample mean	Simple sample covariance
CAPM	Sample mean	CAPM model
Principal	Sample mean	Principal Component Analysis model
Mahalanobis	Sample mean	Shrinkage using Mahalanobis norm
Frobenius	Sample mean	Shrinkage using Frobenius norm
CCC-GARCH	Sample mean	Constant Conditional Correlation GARCH
DCC-GARCH	Sample mean	Dynamic Conditional Correlation GARCH
Market*	N/A	N/A

* *Assign asset weight to each industry sector proportional to its market value:* $w_{ti} = M_{ti} / \sum_{i=1}^{N} M_{ti}$

For every estimator, we use the following portfolio rebalancing strategy: estimate the industry sector weights using the most recent 100 daily returns (except GARCH models, which automatically incorporate exponential weighting in the models) and rebalance the portfolio weights every five trading days, which is usually one business week. Since there are 2600 trading days in the data, there are 500 rebalances in total. In practice, there are transaction costs involved for asset turnovers when we change the weights of each asset using updated information. We will compare the results both without considering transaction costs and with constant 0.05% transaction costs (for each dollar bought or sold, 0.05 cents are paid as trading costs. This estimate is toward the lower end of the trading costs used in the financial literature). We apply target return constraint and convexity constraint to all estimates:

$$w'\mu = \mu_p, w'e = 1$$

For V, CAPM, and principal component methods, we also apply a box constraint that every estimated weight must be in the interval [-1, 1]. (Final calculated weights indicate this restriction makes little difference since, in most cases, industry sector weights are within [-1, 1] without imposing the interval constraints.) The resulting stream of ex-post portfolio returns is collected for each estimator/target return combination. We calculate the following statistics of the ex-post returns of each estimator/target return combination:

Mean: the sample mean of weekly ex-post returns;

STD: the sample standard deviation of weekly ex-post returns;

Information ratio: $IR = mean / STD \times \sqrt{260/5}$, where the standardization by $\sqrt{260/5}$ makes the information ratio an annual estimate assuming 260 trading days per year;

α-VaR for α = 5% and 1%: the loss at the α-quantile of the weekly ex-post return;

α-CVaR for α = 5% and 1%: the sample conditional mean of the losses of the weekly ex-post return distribution, given they are below the α-quantile;

MaxDD: the maximum drawdown, which is the maximum loss in a week;

CRet: cumulative return;

Turnover: weekly asset turnover, defined as mean of absolute weight changes ($\sum_{i=1}^{51}\left|w_{t,i} - w_{t-1,i}\right|$) for 500 updates, where $w_{t,i}$ is the weight of the ith asset at period t.

Cret_cost: cumulative return with transaction costs incorporated.

IRcost: Information ratio with transaction costs incorporated.

The following are the model implementation details:

CAPM: One-factor model using market-value weighted returns as the factor.

Principal: For multiple-factor models, we only test the model using principal components due to our limited access to BARRA factors. Connor and Korajczyk found about six pervasive factors in the cross-section data of New York Stock Exchange and American Stock Exchange stock returns (*19*). In our implementation, instead of fixing the number of principal components, we let the model choose the number of principal components which explains at least 80% of the sample variance. Generally the number of principal components selected (7 – 10) is of similar range to Connor and Korajczyk's study.

Mahalanobis: The shrinkage model combines the single-factor covariance matrix estimated in **CAPM** with the covariance matrix estimated from simple sample covariance matrix. The sum of Mahalanobis distance is used as the cost function to choose the optimal weight λ. The λ is updated for each portfolio rebalance.

Frobenius: Ledoit *et al.* used a simple average return of all asset classes as the market return since they believe it is often better at explaining stock market variance than value-weighted indices. For our application, we implemented the model using exactly the same approach. The prior is the single-factor covariance matrix estimated using simple-average returns. The α is updated for each portfolio rebalance. (A similar approach using a value-weighted market model is also tested and the final difference in portfolio statistics is small.)

Figure 2.1 shows how the estimate of optimal shrinkage intensity evolves through the 10 years in our sample. The shrinkage factor based on the sum of Mahalanobis distances is very stable around 45% over time. It indicates that there is as much estimation error in the sample covariance matrix as there is bias in the CAPM model based on Mahalanobis distance. On the contrary, the shrinkage factors based on the asymptotic Frobenius norm are highly volatile over time. One explanation is that since T is only twice as large as N, the asymptotic estimation based on the Central Limit Theorem may not be stable.

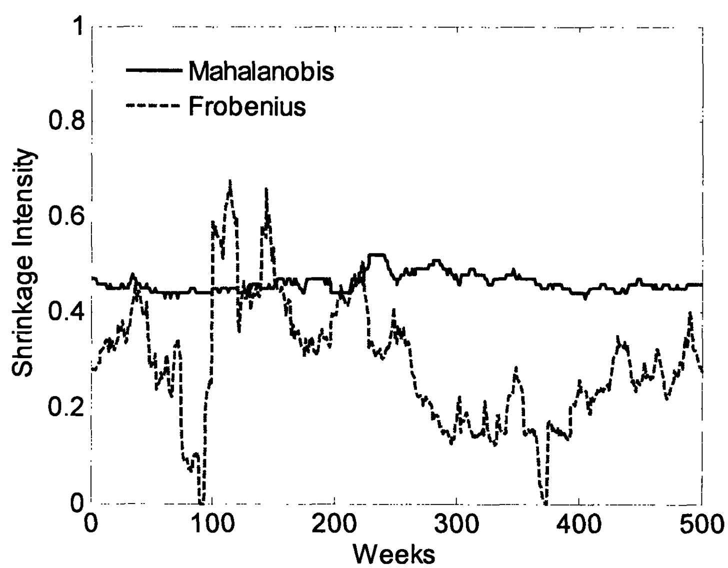

Figure 2.1. Optimal shrinkage intensity estimated using either Mahalanobis distance or asymptotic Frobenius norm estimation

CCC-GARCH: Many of the papers simply used in-sample results to compare different GARCH models. We believe it is more appropriate to use (rolling) out-of-sample validation, especially considering we are applying an executable rebalancing model. To estimate the daily conditional covariance matrix using constant conditional correlation GARCH, the returns are standardized ($\tilde{R}_{i,t} = R_{i,t} / \sqrt{\sigma_{ii,t}}$) and pair-wise sample correlation coefficients are estimated using standardized returns. The conditional variance of each individual is estimated using univariate GARCH and combined with the correlation matrix to estimate the conditional covariance matrix. For each rebalance, we re-run the GARCH model using up-to-date return information to yield new conditional covariances. Because of the computational efficiency of the CCC-GARCH model, the estimation for 500 covariance matrices is done in 100 minutes.

DCC-GARCH: DCC-GARCH based conditional covariance is estimated directly using a multivariate GARCH library published by Kevin K. Sheppard. For DCC-GARCH, the function implements Engle and Sheppard's DCC GARCH estimator (*20*). For each rebalancing, we again re-run the GARCH model using up-to-date return information to yield new conditional covariances.

Scalar & BEKK GARCH: For a single covariance matrix estimation (T = 2600, N = 51), neither model converges after a week of computation. Considering the number of parameters is larger than 1000, the Matlab optimization toolbox may not have an efficient algorithm to solve the problem.

Except for the Market model, which uses market weights and corresponding market returns, a range of target expected annual portfolio returns from 10% to 50% are used for portfolio construction. Appendix II shows detailed return statistics for V, CAPM, Principal, Mahalanobis, Frobenius, CCC-GARCH, DCC-GARCH models and Market index. Table 2-2 and 2-3 show the summarized results for annual expected returns 15% and 20%:

Table 2.2. Performance of V, CAPM, Principal, Mahalanobis, Frobenius, CCC-GARCH, DCC-GARCH models and Market index for $\mu_p = 15\%$

15%	V	CAPM	Principal	Mahalanobis	Frobenius	CCC-GARCH	DCC-GARCH	Market
mean	0.065%	0.028%	0.056%	0.041%	0.074%	0.103%	0.102%	0.160%
STD	1.962%	1.754%	1.676%	1.590%	1.673%	1.612%	1.611%	2.343%
IR	0.239	0.117	0.240	0.186	0.317	**0.459**	**0.454**	**0.491**
VaR(0.05)	3.06%	2.80%	**2.74%**	2.76%	2.69%	**2.58%**	**2.66%**	4.06%
Var(0.01)	5.78%	5.01%	5.02%	4.96%	5.14%	**4.42%**	**4.47%**	5.28%
CVaR(0.05)	4.44%	4.11%	3.78%	3.89%	3.88%	**3.64%**	**3.65%**	5.08%
Cvar(0.01)	6.80%	6.27%	5.60%	6.06%	5.95%	**5.51%**	**5.52%**	7.11%
MaxDD	-7.48%	-7.53%	**-6.56%**	-7.89%	-7.26%	**-6.15%**	**-6.15%**	-10.01%
Cret	1.256	1.067	1.232	1.152	1.346	1.564	1.556	**1.935**
Cret_cost	0.845	0.982	1.022	1.018	1.147	1.104	1.100	**1.923**
Ircost	-0.054	0.049	0.079	0.074	0.179	0.147	0.143	**0.487**
Turnover	1.59	**0.33**	0.75	0.49	0.64	1.39	1.39	**0.02**

The best results for each statistical evaluation are labeled in bold. Turnover above 100% is possible because of short sales.

Table 2.3. Performance of V, CAPM, Principal, Mahalanobis, Frobenius, CCC-GARCH, DCC-GARCH models and Market index for $\mu_p = 20\%$

20%	V	CAPM	Principal	Mahalanobis	Frobenius	CCC-GARCH	DCC-GARCH	Market
mean	0.060%	0.028%	0.051%	0.037%	0.069%	0.099%	0.098%	0.160%
STD	1.978%	1.761%	1.686%	1.596%	1.681%	1.620%	1.619%	2.343%
IR	0.217	0.115	0.219	0.165	0.296	**0.440**	0.435	**0.491**
VaR(0.05)	3.09%	2.93%	2.78%	2.73%	2.72%	**2.62%**	**2.65%**	4.06%
Var(0.01)	6.05%	5.01%	5.14%	4.96%	5.13%	**4.61%**	**4.67%**	5.28%
CVaR(0.05)	4.48%	4.15%	3.82%	3.90%	3.92%	**3.66%**	**3.67%**	5.08%
Cvar(0.01)	6.84%	6.28%	5.66%	6.05%	5.90%	**5.50%**	**5.51%**	7.11%
MaxDD	-7.35%	-7.53%	-6.39%	-7.67%	-7.14%	-6.15%	-6.15%	-10.01%
Cret	1.221	1.065	1.203	1.126	1.315	1.535	1.526	**1.935**
Cret_cost	0.816	0.974	0.992	0.990	1.114	1.081	1.076	**1.923**
Ircost	-0.077	0.042	0.054	0.049	0.154	0.128	0.123	**0.487**
Turnover	1.61	**0.36**	0.77	0.52	0.66	1.40	1.40	**0.02**

Overall, the results indicate that neither the CAPM model nor the Principal Component model yields positive improvement compared with the simple mean/variance approach. The statistical analysis for different models using $\mu_p = 15\%$ and $\mu_p = 20\%$ convey the same information. Here we use $\mu_p = 15\%$ to interpret the results. The CAPM model has an information ratio of 0.117, which is significantly lower than V's IR 0.239. The negative result of CAPM model is likely due to the large bias introduced by a single-factor model. The multiple-factor model using principal components yields better results than CAPM model by reducing the model bias. Both factor models dramatically reduce the turnovers (0.36 for CAPM and 0.77 for Principal vs. 1.61 for V), which reflects structured factor models' ability to reduce estimation variance and that the models are more stable. As a result, both models offer better cumulative returns and higher IR if trading costs are taken into consideration.

The Frobenius norm based shrinkage model introduced by Ledoit *et al.* does yield much better results than Mahalanobis distance. Since $T \approx 2N$, the matrix may be ill-conditioned, so the inverse introduces much calculation error in Mahalanobis distance. On the contrary, the Frobenius norm does not rely on the matrix inverse and is more robust to estimation errors. Compared with the simple V model, the Frobenius model results in higher IR (0.317 vs. 0.239 for V), lower VaR (2.69% vs. 3.06% at α = 5%), lower CVaR (3.88% vs. 4.44% at α = 5%), higher maximum drawdown (-7.26% vs. -7.48%) and much lower turnover rate (0.64 vs. 1.59), all of which indicates that the shrinkage model using the Frobenius norm does improve the performance of mean-variance optimization compared with the simple sample mean-variance approach.

Even though Frobenius model is the best among the factor and shrinkage models, in most statistical aspects it is inferior to the market index. For example, the market index has an IR of 0.491, which is significantly higher than all factor and shrinkage models. Since the US equity market is the most efficient market among all financial markets and instruments, it indicates that the index fund is near the 'true' optimal portfolio and it could be difficult to outperform the market using pure statistical models. The large difference also reflects an important issue that these models fail to address, the influence of outliers. Most of the published market betas, including the one we implemented in Chapter 2, rely on ordinary least squares (OLS) regression. Similar to covariance estimation, the betas estimated from OLS regression are heavily influenced by outliers as well. There is often a large difference between OLS and robust beta due to the outlier-induced distribution of the OLS beta. Similarly, the principal component analysis is also influenced by outliers. As a result, neither the factor models nor the shrinkage models yield a robust estimation for the covariance matrix considering the number of assets is of the same order as the number of daily data points.

GARCH models clearly showed their advantage over the shrinkage models implemented in this Chapter. As shown in Tables 2.2 and 2.3, GARCH models yield IRs slightly lower than market index with lower downside risk as measured by VaR and CVaR. The results also show that constant conditional correlation GARCH and dynamic conditional correlation GARCH have results similar to each other. Although the IR ratio is slightly lower than the market index, the discrepancy may be due to the estimation error in expected returns since the simple mean of historical returns were used. As discussed in Chapter 1, the global minimum variance portfolio (GMVP) does not require an estimation of expected returns μ, so here we use it to independently test the covariance matrix estimation. As shown in table 2.4, GMVP has higher IR compared with either $\mu_p = 15\%$ or $\mu_p = 20\%$ for every model, which indicates the lower IR when μ_p is fixed is partly contributed to the estimation error in expected returns. Both GARCH models yield the overall best results among all models. They both have higher IRs than the market index, although the increase is not large.

Table 2.4. GMVP performance of V, CAPM, Principal, Mahalanobis, Frobenius, CCC-GARCH, DCC-GARCH models and Market index

GMVP	V	Principal	Mahalanobis	Frobenius	CCC-GARCH	DCC-GARCH	Market
mean	0.094%	0.082%	0.066%	0.091%	0.113%	0.113%	0.160%
STD	1.958%	1.697%	1.621%	1.694%	1.639%	1.639%	2.343%
IR	0.345	0.348	0.293	0.387	**0.499**	**0.497**	0.491
VaR(0.05)	3.05%	2.72%	**2.52%**	2.60%	**2.58%**	**2.58%**	4.06%
Var(0.01)	5.78%	4.76%	4.96%	5.14%	**4.70%**	**4.71%**	5.28%
CVaR(0.05)	4.42%	3.84%	3.86%	3.80%	**3.65%**	**3.66%**	5.08%
Cvar(0.01)	6.76%	**5.54%**	6.14%	6.08%	**5.62%**	**5.62%**	7.11%
MaxDD	-8.12%	-6.57%	-7.89%	-7.26%	**-6.15%**	**-6.15%**	-10.01%
Cret	1.451	1.401	1.301	1.466	1.648	1.645	**1.935**
Cret_cost	0.981	1.179	1.164	1.261	1.169	1.168	**1.923**
Ircost	0.056	0.201	0.194	0.259	0.196	0.196	**0.487**
Turnover	1.57	0.69	0.44	0.60	1.38	1.37	**0.02**

The best results for each statistical evaluation are labeled in bold. Turnover above 100% is possible because of short sales.

2.8. Conclusion

Expected returns are more often estimated using a financial/economical/political forecast of different factors in practice instead of relying on data, while the estimation of the covariance matrix is mainly based on financial econometrics where an appropriate model adds value. In the investment management community, arguments are sometimes made in favor of focusing on estimating expected returns alone since many believe that portfolio weights are more sensitive to changes in the expected returns than changes in the covariance matrix. Nevertheless, more and more of the recent results showed the importance of covariance matrix estimation as well.

Using the same sample mean as expected return, we showed that the information ratio is widely different (from 0.117 to 0.459 at $\mu_p = 15\%$) for different covariance estimators, which confirms the importance of covariance matrix estimation. All these results showed that even the well-founded financial econometrical models such as multivariate GARCH either do not outperform the market index or only provide a small performance increase, which indicates the efficiency of the market and the difficulty to consistently outperform the market. However, some of the more sophisticated econometrical approaches also point to the possibility of outperforming the market index if we can more accurately estimate expected returns and covariance and apply the mean-variance optimization in a more robust way, which we will further investigate in the following Chapters.

Chapter 3

Robust Estimation of the Mean and Covariance Matrix

3.1. Introduction

In the previous Chapter, we discussed the simple sample mean and covariance matrix, factor models and shrinkage methods combining these two approaches. The shrinkage methods are designed to reach the right balance between estimation error and estimation bias. However the results from all these models are unsatisfactory when applied to our study. A careful analysis indicates that the problem may lie in the undue influence of outliers. Factor models used in the financial literature usually rely on simple linear regression to estimate the regression coefficients. It is well known in the statistics literature that outliers generated by heavy-tailed or contaminated distributions, which is the case for asset returns, often have a substantial distorting influence on least squares estimates. A small percentage of outliers, in some cases even a single outlier, can distort the final estimated variance and covariance. Evidence has shown that the most extreme (large positive or negative) coefficients in the estimated covariance matrix often contain the largest error and as a result, mean-variance optimization based on such a matrix routinely gives the heaviest weights (either positive or negative) to those coefficient that are most unreliable. This "error-maximization" phenomenon (named by Michaud (*21*)) causes the mean-variance technique to behave very badly unless such errors are corrected.

It is therefore important to use robust estimators that are less influenced by outliers. This issue has been extensively researched in the statistical literature, with regard to both theory and applications. During the past three decades, statisticians have developed a variety of robust estimation methods to estimate both the mean (location) and the covariance matrix (scatter matrix) (*22-25*). However, the use of robust estimators has received relatively little attention in the finance literature overall, and in the context of estimating the expected value and the covariance matrix of asset returns in particular (*26, 27*). There are several factors contributing to this phenomenon. First, some of the early robust approaches (e.g. least absolute deviations) yield either negative results or somewhat unimpressive positive results, so these approaches attracted little attention in practice. Second, to a certain extent, including longer historical data mitigates the outlier problem since each data point has less influence on the final estimation. Last but not least, many of the more complicated robust methods have been developed or put into practice only in recent years due to the intensive computation involved in contrast to simple mean and covariance matrix methods. These methods have been developed in academia and applied mainly to robust linear regression. There is often a considerable time lag before financial practitioners utilize these methods. In this Chapter, we will take the initiative to investigate the value of these approaches in finance, especially in asset allocation problems.

3.2. Traditional Robust Estimators

3.2.1. LAD portfolio estimator

Among all the robust estimation methods, the least absolute deviation (LAD) estimator is one of the oldest and most widely known robust alternatives to least squares. The LAD method was first applied to financial studies by Sharpe in 1971 to estimate market beta (*28*). Instead of minimizing the sum of the squared deviations (LSE: $\min \sum_i (y_i - \hat{y}_i)^2$, where y_i 's are the observed values and $\hat{y}_i$ is the predicated values by the regression model) for linear regression, Sharpe obtained the regression line by minimizing the sum of absolute deviations ($\min \sum_i | y_i - \hat{y}_i |$). By reducing the quadratic term to a linear one, the LAD approach intuitively reduces the influences of outliers. When applied to individual stocks, the estimated betas are often significantly different from the LSE estimates. The LAD method can also be used as an alternative approach to portfolio selection based on a simple observation by Lauprete (*29*). Again, if the expected return and covariance matrix are estimated from the historical sample $R_{1\bullet}, \cdots, R_{T\bullet}$, the original optimization problem

$$\min_{\varpi} \tfrac{1}{2} w' \Sigma w \quad s.t.\, w'\mu = \mu_p,\, w'e = 1 \tag{3.1}$$

can be rewritten as

$$\min_{w,q} \tfrac{1}{T} \sum_{t=1}^{T} (w' R_t - q)^2 \quad s.t.\, w'\mu = \mu_p,\, w'e = 1. \tag{3.2}$$

If we replace the objective function's quadratic term with absolute deviation, an alternative portfolio optimization problem is formulated as

$$\min_{w,q} \tfrac{1}{T} \sum_{t=1}^{T} \left| w' R_t - q \right| \quad s.t.\, w'\mu = \mu_p,\, w'e = 1. \tag{3.3}$$

LAD portfolio optimization is actually a specific example of the shortfall risk (as defined in section 1.4) with $\alpha = 0.5$:

$$s_{0.5}(w) = w'\mu - E[w'r \mid w'r \le q_{0.5}(w)], \tag{3.4}$$

where r is a $N \times 1$ column vector of r_is with $r_i, \forall i = 1, \cdots, N$ representing the return of the ith asset in the portfolio; $q_{0.5}(w) = \inf\{z \mid P(w'r \le z) \ge 0.5\}$ is the 50%-quantile of the portfolio. In general, the portfolio that minimizes α-shortfall risk can be expressed as

$$\min_{w} \min_{q} \tfrac{1}{T} \sum_{t=1}^{T} \rho_\alpha (w' R_t - q) \quad s.t.\ w'\mu = \mu_p,\ w'e = 1 \tag{3.5}$$

where $\rho_\alpha(z) = z - \frac{1}{\alpha} z I_{\{z<0\}}$ with I as an indicator function (*29*).

3.2.2. Huber portfolio estimator

The LSE approach minimizes the sum of squared deviations, which is highly sensitive to outliers because of the use of the square. The LAD approach minimizes the sum of absolute deviations, which is robust but unfortunately not very efficient. Huber (1964) proposed a compromise between these two using the Huber loss function (*30*) with parameter γ:

$$\eta_\gamma(z) = \begin{cases} z^2 & if\ |z| \le \gamma \\ 2\gamma |z| - \gamma^2 & if\ |z| > \gamma \end{cases} \tag{3.6}$$

The corresponding Huber portfolio estimator can be expressed as:

$$\min_{w,q} \tfrac{1}{T} \sum_{t=1}^{T} \eta_\gamma (w' R_t - q) \quad s.t.\ w'\mu = \mu_p,\ w'e = 1. \tag{3.7}$$

In this method, deviations beyond $+/-\ \gamma$ of the location parameter are penalized linearly instead of quadratically, so it is less sensitive to outliers than variance portfolio estimators. This adjustment could be valuable considering that asset returns usually have heavy tails. The parameter γ is often chosen as proportional to the mean absolute deviation (MAD), which is defined as

$$MAD = E[| r - q_{0.5} |] \tag{3.8}$$

where $q_{0.5}$ is the median of r's distribution. Using historical samples, the sample MAD can be estimated as

$$MAD(\hat{w}_{0.5}) = \tfrac{1}{T}\sum_{t=1}^{T}\left|\hat{w}_{0.5}{}'R_t - \hat{q}_{0.5}\right|, \tag{3.9}$$

where $\left(\hat{w}_{0.5}, \hat{q}_{0.5}\right)$ are estimated using the LAD method as discussed in the section 3.2.1. Then we can set

$$\gamma = z_{1-\alpha}\sqrt{\tfrac{\pi}{2}}MAD(\hat{w}_{0.5}) \tag{3.10}$$

where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal.

3.2.3. Rank-correlation portfolio estimator

For any variable pair X and Y, the simple sample correlation is also known as the Pearson correlation, which is based on the assumption that both X and Y values are sampled from a joint-normal distribution, at least approximately. An alternative correlation estimator, the nonparametric Spearman correlation is based on the rankings of the two variables, and so makes no assumption about the distribution of the values. As a result, it has the apparent advantage of not being influenced by outliers. To our knowledge, the Spearman correlation has been broadly used by many quantitative researchers, including the research groups in some of the world's largest investment firms such as State Street Global Advisors. Nevertheless, the efficiency of this method can be low since it misses some of the valuable information incorporated in parametric analysis. Combining Spearman rank correlation and estimation of standard deviations for each variable, we can estimate the covariance matrix as well. For asset allocation problem, the correlation $\hat{\rho}_{ij}$ of each pair of asset $R_{\bullet i}$ and $R_{\bullet j}$ can be estimated using Spearman rank-order correlation. Then we can estimate the pair-wise covariance using $\hat{\sigma}_{ij} = \hat{\rho}_{ij} s_i s_j$ (where s_i is the standard deviation of asset *i*) and construct the covariance matrix $\hat{\Sigma} = [\hat{\sigma}_{ij}],\ \ i, j = 1, \cdots, N$.

Interestingly, simple sample standard deviation is often used in practice to estimate the final covariance matrix. Such estimation is again influenced by outliers and it is also not consistent with nonparametric ideas behind the Spearman correlation. In the presence of outliers, we believe it is more appropriate to use robust estimators to estimate each asset's expected return and standard deviation. The sample median is a natural choice for the expected return estimation and the scaled interquartile range (IQR) can be used as a robust unbiased estimate of standard deviation respectively for each variable (column):

$$m_i = median(X_{\bullet i}),\ s_i = 0.7413 \times IQR(X_{\bullet i}). \tag{3.11}$$

An alternative measure of scatter uses adjusted median absolute deviation from the median:

$$s_i = \frac{MAD_i}{0.6745} = \frac{median(|X_{\bullet i} - m_i|)}{0.6745}. \tag{3.12}$$

For both methods, adjustment factors are used to yield unbiased estimators if the sample is drawn from a normal distribution. Considering the value of the GARCH models, we can also estimate the conditional variance of each asset using univariate GARCH.

3.3. M-estimator

A well-known class of robust location and scatter estimators, M-estimators, was introduced by Maronna (*22*) and has been extensively studied (*31*). The robust location vector t and the scatter matrix V are defined as solutions of the system:

$$\begin{aligned} &\frac{1}{n}\sum_i u_1(d_i)(x_i - t) = 0, \\ &\frac{1}{n}\sum_i u_2(d_i^2)(x_i - t)(x_i - t)' = V \end{aligned} \tag{3.13}$$

where x_i s are cross-sectional observations (row vectors); $d_i^2 = (x_i - t)V^{-1}(x_i - t)'$, and u_1 and u_2 are non-increasing functions satisfying a set of general assumptions (*22*).

More intuitively, the robust estimators of means (a row vector) and covariance are expressed as

$$\begin{aligned} &\mu = \sum_i z_i x_i / \sum_i v_i \\ &V = \sum_i z_i^2 (x_i - \overline{x})(x_i - \overline{x})' / \sum_i z_i^2 \end{aligned} \tag{3.14}$$

where $d_i = \sqrt{(x_i - \overline{x})V^{-1}(x_i - \overline{x})'}$ (*31, 32*) and $z_i = f(d_i)/d_i$ with $f()$ as a function of d_i.

The general principle is to give full weight to observations assumed to come from the main body of the data, but reduce weight for the observations from tails of the contaminated data. Simple mean/variance estimation is a special example of an M-estimator with unit weight for all data points. The Huber estimator discussed in the last section is another example of an M-estimator, which has a monotone decreasing weight γ / d_i if $d_i > \gamma$ and 1 otherwise (the function $f(d_i) = \gamma$ if $d_i > \gamma$ and $f(d_i) = d_i$ otherwise).

The M-estimates for the covariance matrix are positive semidefinite and affine equivariant, which means the estimator behaves properly under affine transformations of the data. That is, for a data set X (a $T \times N$ matrix) the estimate $(\hat{\mu}, \hat{\Sigma})$ satisfies

$$\hat{\mu}(XA + ev') = \hat{\mu}(X)A + v' \text{ and } \hat{\Sigma}(XA + ev') = A'\hat{\Sigma}(X)A \qquad (3.15)$$

where A is a $N \times M$ matrix; e is a $T \times 1$ vector of ones; v is a $M \times 1$ vector;

Many of the early M-estimators for multivariate location and scatter have low breakdown points, the maximum proportion of outliers that the estimate can safely tolerate is on the order of $T/(N+1)$. So for high-dimensional problems, as in the case of asset allocation, robust methods with higher breakdown points are often desirable.

3.4. Fast-MCD

To increase the breakdown point, the minimum volume ellipsoid method (MVE) and the minimum covariance determinant (MCD) (*33*) mthods, introduced by Rousseeuw (*34*), look for the ellipsoid with smallest volume that covers h data points and smallest determinant of the covariance matrix for h data points respectively, where $T/2 \le h < T$. Both MVE and MCD use the average and the covariance matrix of those identified h points to estimate the population mean and the covariance matrix and both have a break-down value of $(T-h)/T$. If the data come from a multivariate normal distribution, the average of the the optimal subset is an unbiased estimator of population mean, yet a finite sample correction factor is required to adjust the covariance matrix to make the covariance unbiased: $\hat{\mu}_{MCD} = \hat{\mu}_{opt}, \hat{\Sigma}_{MCD} = c_{h,n}\hat{\Sigma}_{opt}, c_{h,n} \ge 1$. The multiplication factor $c_{h,n}$ can be determined through Monte Carlo simulation. For our specific purpose, the bias by itself does not affect the weight allocation since all pairs of covariances were underestimated by the same factor. MCD is believed to be a better approach compared with MVE since it is more efficient, more precise than MVE, and is better suited to identify and eliminate the effects of multivariate outliers (*35, 36*).

Until 1999, MCD had rarely been applied to high-dimensional problems because it was extremely difficult to compute. MCD estimators are solutions to highly non-convex optimization problems, which have exponential complexity of the order 2^N in terms of the dimension N of the data. They also have a quadratic computational complexity in the number of observations. So these original methods are not suitable for asset allocation problems when $N > 20$. In order to cope with computational complexity problems, a variety of heuristic methods were proposed to yield "good" estimates with little sacrifice in accuracy. One of the methods, the FAST-MCD algorithm developed by Rousseeuw and Van Diressen (*37*), offers just such an efficient robust estimator. A naïve MCD approach would compare the MCD up to $\binom{T}{h}$ subsets, while FAST-MCD uses sampling to reduce the computation and usually offers a satisfactory heuristic estimation.

The key step of the FAST-MCD algorithm takes advantage of the fact that, starting from any approximation to the MCD, it is possible to compute another approximation with a determinant no higher than the current one. The method is based on the following theorem related to a concentration step (C-step):

Let $H_1 \subset \{1, \cdots, n\}$ be any h-subset of the original cross-sectional data, put $\hat{\mu}_1 = \frac{1}{h}\sum_{j \in H_1} x_j$ and $\hat{\Sigma}_1 = \frac{1}{h}\sum_{j \in H_1} (x_j - \hat{\mu}_1)(x_j - \hat{\mu}_1)'$. If $\det(\hat{\Sigma}_1) \neq 0$ define the distance $d_1(j) = \sqrt{(x_j - \hat{\mu}_1)\hat{\Sigma}_1^{-1}(x_j - \hat{\mu}_1)'}$,

$j = 1, \cdots, T$ Now take H_2 such that $\{d_1(i); i \in H_2\} := \{(d_1)_{1:T}, \ldots, (d_1)_{h:T}\}$ where $(d_1)_{1:T} \le (d_1)_{2:T} \le \ldots \le (d_1)_{n:T}$ are the ordered distances, and compute $\hat{\mu}_2$ and $\hat{\Sigma}_2$ based on H_2. Then $\det(\hat{\Sigma}_2) \le \det(\hat{\Sigma}_1)$ with equality if and only if $\hat{\mu}_2 = \hat{\mu}_1$ and $\hat{\Sigma}_2 = \hat{\Sigma}_1$.

If $\det(\hat{\Sigma}_1) > 0$, the C-step yields $\hat{\Sigma}_2$ with $\det(\hat{\Sigma}_2) \le \det(\hat{\Sigma}_1)$. Basically the theorem indicates the sequence of determinants obtained through C-steps converge in a finite number of steps from any original *h*-subset to a subset satisfying $\det(\hat{\Sigma}_{m+1}) = \det(\hat{\Sigma}_m)$. Afterward, running the C-step no longer reduces the determinant. However, this process only guarantees that the resulted $\det(\hat{\Sigma})$ is a local minimum instead of the global one. To yield the *h*-subset with global minimum $\det(\hat{\Sigma})$ or at least close to optimal, many initial choices (often > 500) of H_1 are taken and C-steps are applied to each. To reduce the amount of computation and the influence of potential outliers, the method also implements the following strategies:

1. To yield good initial subset which has lower probability of including outliers, a random $(N+1)$-subset *J*, instead of *h*-subset, is drawn and extended to *h*-subset by the following procedures: calculate $\hat{\mu}_0$ and $\hat{\Sigma}_0$ using subject *J*; compute the distance $d_0(j) = \sqrt{(x_j - \hat{\mu}_0)\hat{\Sigma}_0^{-1}(x_j - \hat{\mu}_0)'}$ and sort the distance into $d_0(\pi(1)) \le \cdots \le d_0(\pi(T))$; select $H_1 = \{\pi(1), \cdots, \pi(h)\}$ as the starting *h*-subset. This strategy may seem cumbersome, yet it reduces the probability of starting with a bad subset H_1. When starting from a bad subset, the iterations will not converge to a useful solution. If the percentage of outliers is high, the probability of a $(N+1)$-subset without outliers is much higher than an h-subset and the final result will generally be more robust.

2. Since each C-step involves the calculation of covariance matrix, its determinant and Mahalanobis distances, an extra strategy is implemented to reduce the number of C-steps. It is observed that after two C-steps, many runs that will lead to a global minimum already have had considerably smaller determinants. So instead of carrying on C-steps for each initial subset until a local minimum is reached, two C-steps are applied to each initial subset and further C-steps are only applied to 10 subsets with lowest determinants.

3. For large T ($T > 600$), the C-steps are often carried out in several nested random subsets, starting with small subset around 300 observations and ending with the entire dataset of T observations. For our study, T is only 100, so this strategy is not implemented in estimation.

Simulated and empirical results showed that FAST-MCD typically gives "good" results and is orders of magnitude faster than MCD methods. Yet, the FAST-MCD method still requires substantial running times for large N and T, and the probability of retaining outliers in the final h-subset increases when T becomes large. To our knowledge, no financial literature has applied FAST-MCD method in asset returns' expected value and covariance matrix estimation. So in this thesis, we will implement the FAST-MCD methods to investigate its potential in asset allocation.

3.5. Pair-wise Robust Estimation

Similar to the original MCD approach, FAST-MCD is an affine equivariant estimator. If the affine equivariance requirement is dropped, much faster estimators with high breakdown points can be computed. These methods are often based on pair-wise robust correlation or covariance estimates such as coordinate-wise outlier insensitive transformations (e.g. Huber-function transformation, quadrant correlation) and bivariate outlier resistant models. These pair-wise approaches can have high breakdown points for the overall covariance. One example of pair-wise robust estimation is the Spearman correlation we discussed in section 3.3. All these methods have quadratic complexity in the number of variables and linear complexity in the number of observations, so they reduce the computational complexity from $O(2^N T^2)$ to $O(N^2 T)$. We can either directly estimate robust pair-wise covariance, or estimate robust standard deviation and correlation separately or sequentially and combine them to estimate pair-wise covariance using the formula $\hat{\sigma}_{ij} = \hat{\rho}_{ij}\hat{\sigma}_i\hat{\sigma}_j$.

Besides the computational efficiency, the pair-wise model assumes a fundamentally different contamination model for data that contain outliers. MCD models assume the majority of the data come from a multivariate normal distribution and the remainder come from a different distribution. So the data are from the following mixed model:

$$F = (1-\varepsilon)F_0 + \varepsilon H \quad 0 < \varepsilon < \tfrac{1}{2} \tag{3.16}$$

where F is the mixed model; F_0 is a multivariate normal distribution; H is an arbitrary multivariate distribution that generates outliers.

Basically MCD models assume that each row is either from the core distribution F_0 or outlier generating distribution *H*. Such a contamination model is rather restrictive for our application. By looking at *N*-dimensional outliers, the models basically assume that all asset returns for any given day are either from a core distribution F_0 or outlier generating distribution *H*. This assumption is only true if the market is the only factor that determines asset returns or there are high correlations between different assets' returns. In practice the market return by itself only explains a small percentage of variance of asset returns. Industrial factors and idiosyncratic risk have been shown to explain the majority of the return variances. The pair-wise models use a much more flexible mixed model for data:

$$F = (I - E)F_0 + EH \ \ with \ \ E = diag([\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_N]) \tag{3.17}$$

where we can assume any format for the correlation matrix of $[\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_N]$. MCD models assume complete dependence $\varepsilon_1 = \varepsilon_2 = \cdots = \varepsilon_N$, while pair-wise models often assume independent ε_i and ε_j, $i \neq j$ or independently evaluate the correlation for each pair of ε_i and ε_j. Pair-wise robust methods have mainly been applied to robust linear regression. For asset allocation robust mean and covariance matrix estimation problems, these methods often suffer the drawback of failing to yield a semi-positive definite covariance matrix. In *2002* Maronna *et al.* proposed a good method for obtaining a positive definite covariance matrix, which we will discuss in detail in implementation of quadrant correlation method (*38*). For this specific study, we will adopt and extend two recently-developed pair-wise robust covariance matrix estimation methods to show the value of these new robust estimators in asset allocation.

3.5.1. Quadrant correlation method

The quadrant correlation method, reported by Alqallaf et al (*39*), is a method using pair-wise quadrant correlation coefficients as basic building blocks that includes three steps:

Step A. Compute simple robust location and scale estimates for each column using median and adjusted IQR/MAD.

Step B. Compute biased-adjusted quadrant correlation estimates and initial robust covariance matrix. First calculate the quadrant correlation for each pair as

$$\hat{r}_{ij} = \frac{\sum_{t=1}^{N} sign(y_{ti}) sign(y_{tj})}{n_{ij,0}}, \tag{3.18}$$

where $y_{ti} = x_{ti} - m_i$ and $n_{ij,0}$ is the number of rows such that neither y_{ti} nor y_{tj} is zero. When x_{ti} and x_{tj} are jointly normal with correlation ρ_{ij}, the value r_{ij} of $\hat{r}_{ij}$ in large samples satisfies $|r_{ij}| \le |\rho_{ij}|$, with strict inequality except in the trivial cases $|\rho_{ij}| = \pm 1$. As a result, the estimator r_{ij} is intrinsically biased. The appropriate transformation function for QC estimator turns out to be Kendall's Tau transform:

$$g_{QC}(r) = \sin((\pi/2)r), \tag{3.19}$$

which will yield an unbiased estimator at a Gaussian model. So we can compute the "bias-corrected' QC as Kendall's tau Transform $\rho_{ij} = \sin\left(\frac{\pi}{2}\hat{r}_{ij}\right)$, derive the pair wise robust covariance estimates $c_{ij} = s_i s_j \rho_{ij}$ and the initial covariance matrix $\hat{C}_0 = \{c_{ij}\}$.

Step C. Form the final positive definite robust covariance matrix. Since $\hat{C}_0$ may not be positive semidefinite, adjustment using the method developed by Maronna *et al.* (ref. 38) is applied to guarantee the positive semidefiniteness of the final matrix. Any positive semidefinite covariance matrix can be expressed as $C = \sum \hat{\lambda}_i \hat{a}_i \hat{a}_i'$, where $0 \le \hat{\lambda}_1 \le \cdots \le \hat{\lambda}_N$ are the eigenvalues and $\hat{a}_i\ (i = 1, \cdots, N)$ are the corresponding eigenvectors. If C is not positive semidefinite, then one or some of the eigvalues are negative. To convert such a matrix to a positive semidefintie one, a natural approach is to replace these negative eigenvalues to positive ones. When C is the sample correlation, $\hat{\lambda}_i$'s are the variances of the projected data on the direction of the corresponding eigenvectors. This indicates that in order to get rid of possibly negative eigenvalues in the quadrant covariance matrix $\hat{C}_0$, one can replace the $\hat{\lambda}_i$'s in $C_0 = \sum \hat{\lambda}_i \hat{a}_i \hat{a}_i'$ by the square of robust scale estimates for the projected data. We can compute the decomposition of $\hat{C}_0$: $\hat{C}_0 = Q \Lambda Q'$, where Q is the orthogonal matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues. Then we can transform X to $\tilde{X}$ using the new basis Q: $\tilde{X} = XQ'$ and compute the robust scale estimate ($\tilde{s}_j$) of the columns of $\tilde{X}$ as in step A. Let $\tilde{D}$ be the diagonal matrix whose elements are $\tilde{s}_j^2$ ordered from largest to smallest. The final positive definite robust covariance matrix is $\hat{\Sigma} = Q\tilde{D}Q'$.

Similar to Spearman correlation, the traditional quadrant correlation matrix (as shown in step B) is an nonparametric robust estimator. They were both widely used in the applied sciences to estimate correlation between different variables. Yet unlike the Spearman correlation, the Quadrant correlation matrix is not necessarily positive semidefinite because of Kendall's Tau Transform. For the mean-variance optimization problem, it is necessary to have a positive semidefinite covariance matrix to avoid negative portfolio variance. Since Maronna's method was only published in recent years, we have found no published article applying quadrant covariance matrix to asset allocation problems.

3.5.2. 2D-Winsorization method

Huber's function, defined as $\psi_c(x) = \min\{\max\{-c, x\}, c\}, c > 0$, is also widely used for one-dimensional Winsorization approach. For each of the univariate observations $x_i, i = 1, \cdots, N$, the transformation $u_i = \psi_c((x_i - m_i)/s_i$ is used to shrink the outliers towards the median. Basically it brings the outliers of each variable to the boundary $m_i \pm c \times s_i$ and as a result reduces the impact of outliers. The one-dimensional Winsorization approach is a popular method in finance because of its intuitive appeal and easy computation. Yet for covariance analysis, the method fails to take the orientation of the bivariate data into consideration. To address the problem, Khan *et al.* proposed a bivariate Winsorization method in *2005* (*40*). For each pair of variables, outliers are shrunken to the border of an ellipse which includes the majority of the data by using the bivariate transformation $\tilde{x}_t = \mu_0 + \min(\sqrt{c / D(x_t)}, 1)(x_t - \mu_0)$ with $x_t = [x_{1t}\, x_{2t}]$. Here $D(x)$ is the Mahalanobis distance based on initial bivariate covariance matrix Σ_0 and location μ_0 (estimated using the median), and c is a constant. The proposed method used an adjusted one-dimensional Winsorization method for the initial covariance matrix estimation, calculated the Mahalanobis distance based on this the initial covariance matrix Σ_0 and then applied the transformation to shrink the outliers. Since the paper did not discuss the implementation of the method in detail and did not address the issue of guaranteeing positive definiteness of the covariance matrix, we construct a 2D-Winsorization method combining the Winsorization ideas from Khan's paper and Maronna's method to guarantee the positive semi-definiteness of the covariance matrix:

Step A. Initial covariance estimate. For each pair of variables x_i, x_j, first compute simple robust location (median) and scale (adjusted MAD) estimates for each variable. Following Khan's idea, we use an adjusted Winsorization method that is more resistant to bivariate outliers. For simple one-dimensional Winsorization method, only one tuning constant c_0 is used to shrink all variable pairs (x_i, x_j) within a rectangle with the boundary $m_i \pm c_0 \times s_i$ and $m_j \pm c_0 \times s_j$ In the adjusted Winsorization method, two tuning parameters are used instead with c_1 for the two quadrants (separated by m_i and m_j) that contain the majority of the data and a smaller constant c_2 for the other two quadrants. For example, c_1 can be taken to be 1.96 ($\mu \pm 1.96\sigma$ includes

95% of the data from the normal distribution) and $c_2 = hc_1$ where $h = n_2 / n_1$ with n_1 the number of observations in the major quadrants and $n_2 = T - n_1$, where T is the total number of observations. As shown in Figure 3.1, the data are now shrunk to the boundary of the four smaller rectangles instead of a large rectangle. As a result, the adjusted Winsorization method handles bivariate outliers better than the univariate Winsorization. However, it does raise a problem that the initial covariance matrix constructed from pair-wise covariance may not be positive definite. To address the problem, Maronna's transformation is applied to covert the initial covariance matrix Σ_0 to a positive definite one.

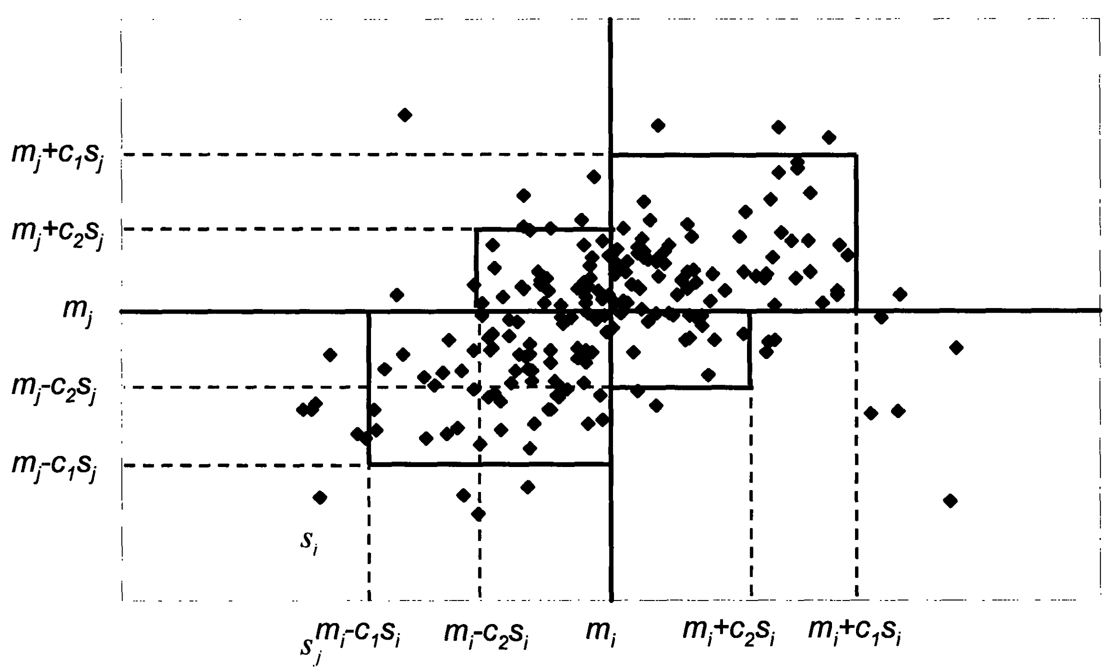

Figure 3.1. Adjusted Winsorization (for initial covariance) with $c_1 = 1.96$, where s_i and s_j are estimated from adjusted MAD

Step B. 2D-Winsorization based covariance matrix. For each pair of (x_i, x_j), outliers are shrunk to the border of an ellipsoid by using the transformation $\tilde{x}_t = \mu_0 + \min(\sqrt{c / D(x_t)}, 1)(x_t - \mu_0)$, with constant c = 5.99 (the 95% quantile of the χ_2^2 distribution). The covariance for each pair is calculated using modified data. Maronna's transformation is again applied to guarantee the positive definiteness of the matrix.

3.5.3. 2D-Huber method

Base on the idea of shrinking data toward the border of a two-dimensional ellipse, we also apply the following approach based on the idea of 2D-Huber transformation (*41*):

Step A. For each pair of variable x_i, x_j, compute simple robust location and scale estimates for each column and construct the initial estimate of mean and covariance matrix as

$$\mu_0 = \begin{bmatrix} med(x_i) \\ med(x_j) \end{bmatrix}, \Sigma_0 = \begin{bmatrix} \frac{MAD(x_i)}{0.6745} & 0 \\ 0 & \frac{MAD(x_j)}{0.6745} \end{bmatrix} \text{ or } \Sigma_0 = \begin{bmatrix} 0.7413IQR(x_i) & 0 \\ 0 & 0.7413IQR(x_j) \end{bmatrix} \quad (3.20)$$

Step B. For each μ_k, Σ_k, calculate the Mahalanobis distance for each data's return pair

$$D_t = \left(x_{t,ij} - \mu_k\right)' \Sigma_k^{-1} \left(x_{t,ij} - \mu_k\right) \quad (3.21)$$

where $x_{t,ij} = \begin{bmatrix} x_{ti} \\ x_{tj} \end{bmatrix}$ and then calculate the weight for each $x_{t,ij}$ using un-normalized $z_t = \min(\sqrt{c / D(x)}, 1)$. The update μ_{k+1}, Σ_{k+1} are then estimated as

$$\mu_{k+1} = \sum_{i=1}^{T} z_t x_t / \sum_{i=1}^{T} z_t \text{ and } \Sigma_{k+1} = \sum_{i=1}^{T} z_t^2 (x_t - \mu_{k+1})(x_t - \mu_{k+1})' / \sum_{i=1}^{T} z_t^2 . \quad (3.22)$$

Such iteration is repeated until μ_{k+1}, Σ_{k+1} and μ_k, Σ_k converge as determined by the sum of absolute difference between two consecutive Σ is less than a predefined error ε. The covariance variable x_i, x_j is then set as Σ_{k+1}.

Step C. Combine the estimated covariances for all pairs to yield an initial covariance matrix. The final positive definite robust covariance matrix was derived using the same step C as in the quadrant correlation method.

Overall, the computational speed of the 2D-Huber method is significantly slower than quadrant correlation and 2D-Winsorization methods because of iterations involved for each pair of variables; however, it is still faster than the FAST MCD method for the same data set. Compared with the 2D-Winsorization method, the 2D-Huber method is likely to yield a more robust estimation of the covariance because of repeated iteration. However, we expect the difference because of the robust estimation of the initial covariance matrix in 2D-Winsorization method to reduce the effect of bivariate outliers. It is also worth noting that the estimated covariance matrix often slightly underestimates the real covariance, so the estimation is biased. Yet it is believed that for the constant c = 5.99 (the 95% quantile of the χ_2^2 distribution) that we choose, the bias would be small. Furthermore, the asset weights depend on the relative size of the covariance, so the impact of bias on our problem is even smaller.

3.6. Application to Historical Data

Again, the application uses the same daily return data on 51 MSCI industry sector indexes, from 01/03/1994 to 07/03/2005. For every estimator, we use the following portfolio rebalance strategy: estimate the industry sector weights using the most recent 100 daily returns and rebalancing the portfolio weight every 5 trading days. We study the performance of the following estimators:

Table 3.1. Application of robust estimation to expected returns and covariance matrix

Method	Expected Return Estimation	Covariance matrix Estimation
LAD	Mean	N/A
Huber	Mean	N/A
Rankcov	Median	Rank correlation & robust STD
FAST-MCD	FAST-MCD (95%)	FAST-MCD (95%)
QC-IQR	Median	QC-IQR
2D-Winsorization	Median	2D-Winsorization
2D-Huber	Median	Two-dimensional Huber

FAST-MCD (95%) means h/T=0.95

The following are the model implementation details:

LAD: The original formation for the LAD portfolio optimization problem

$\min_{w,q} \frac{1}{T}\sum_{t=1}^{T}|w'R_i - q|$ $s.t.\ w'\mu \geq \mu_p,\ w'e = 1$ can be reformulated as a linear programming problem:

$$
\begin{aligned}
&\min_{w,q,z}\ w'\overline{R} - q + \frac{1}{0.5\times T}\sum_{t=1}^{T} z_t \\
&s.t.\quad w'e = 1,\ w'\mu \geq \mu_p \\
&\qquad\ z_t \geq q - w'R_t,\ t = 1,\cdots,T \\
&\qquad\ z_t \geq 0,\ t = 1,\cdots,T
\end{aligned}
\tag{3.22}
$$

The reformulated LP has $N+1+T$ variables and $M+2T$ constraints, with $N+1$ dense columns. The problem can be efficiently solved by software package such as Matlab Optimization toolbox.

Huber: For the Huber portfolio estimator

$$
\min_{w,q} \frac{1}{T}\sum_{t=1}^{T}\eta_\gamma(w'R_t - q)\quad where\quad \eta_\gamma(z) = \begin{cases} z^2 & if\ |z| \leq \gamma \\ 2\gamma|z| - \gamma^2 & if\ |z| > \gamma \end{cases}
$$

$s.t.\ w'\mu \geq \mu_p,\ w'e = 1$

We directly use $(\hat{w}_{0.5}, \hat{q}_{0.5})$ from the LAD method to estimate $MAD(\hat{w}_{0.5})$ and then we set the threshold $\gamma = z_{1-\alpha}\sqrt{\frac{\pi}{2}}MAD(\hat{w}_{0.5})$. α is set to 0.10 in our study.

The Huber Portfolio Estimator can be formulated using the following quadratic programming problem:

$$\begin{aligned}
&\min_{w,q,t} \tfrac{1}{T}\sum_{t=1}^{T} u_t^2 + \tfrac{1}{T}\sum_{t=1}^{T} 2\gamma(z_t^+ + z_t^-) \\
&s.t. \quad w'e = 1 \\
&\qquad\; w'\bar{R} \geq \mu_p \qquad\qquad\qquad\qquad \Rightarrow \quad -\bar{R}'w \leq -\mu_p \\
&\qquad\; z_t^+ - z_t^- = w'R_t - q - u_t,\ i = 1,\cdots,T \quad \Rightarrow \quad w'R_t - q - u_t - z_t^+ + z_t^- = 0 \\
&\qquad\; z^+, z^- \geq 0
\end{aligned} \tag{3.23}$$

which has $N+1+3T$ variables and $M+3T$ constraints, with $N+1$ dense columns.

Rankcov: The correlation of each asset pair is calculated using Spearman rank correlation and robust standard deviation is estimated as adjusted interquartile range. Then the pair-wise covariance is estimated as $\hat{\sigma}_{ij} = \hat{\rho}_{ij} s_i s_j$.

FAST-MCD: For FAST-MCD covariance matrix estimation, we make use of the fastmcd function included in MATLAB Library for Robust Analysis written by Sabine Verboven and Mia Hubert. Since the original function has the constraints that $T \geq 5N$ and $N \leq 50$, it is modified to accommodate our estimation.

QC-IQR: A Matlab function qciqr is written to estimate the covariance matrix using Quadrant correlation and adjusted interquartile range. The initial covariance matrix is also converted to a semi-positive definite one using Maronna's method.

2D-Winsorization: c_1 is chosen to be 1.96 in the initial covariance matrix estimation and c is chosen to be 5.99 for the 2-dimensional shrinkage.

2D-Huber. A Matlab function is written to estimate the covariance matrix starting from adjusted MAD. The iteration is stopped when $\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\Sigma_{k+1,i,j}-\Sigma_{k,i,j}\right|<5\times10^{-6}$. The initial covariance matrix is also converted to a semi-positive definite one using Maronna's method.

Again a range of target expected portfolio returns from 10% to 50% annual rate are used for portfolio construction. Appendix 3 shows detailed return statistics for LAD, Huber, FAST-MCD, Rankcov, QCIQR, 2D-Winsorization and 2D-Huber. Tables 3.2 – 3.4 show the summarized results for annual expected return 15% and 20% as well as global minimum variance portfolio.

Table 3.2. Performance of V, LAD, Huber, Rnakcov, FAST-MCD, QCIQR, 2D-Winsorization, 2D-Huber models and Market index for $\mu_p = 15\%$

15%	**V**	**LAD**	**Huber**	**Rankcov**	**FAST-MCD**	**QCIQR**	**2D-winsor**	**2D-Huber**	**Market**
mean	0.065%	0.041%	0.054%	0.078%	0.096%	0.136%	0.155%	0.156%	0.160%
STD	1.962%	2.045%	1.970%	1.953%	2.025%	1.959%	2.007%	1.948%	2.343%
IR	0.239	0.143	0.199	0.290	0.341	0.500	0.558	**0.578**	0.491
VaR(0.05)	3.06%	3.48%	3.08%	3.25%	3.10%	**3.10%**	3.23%	**3.10%**	4.06%
Var(0.01)	5.78%	5.22%	5.88%	5.59%	6.33%	5.70%	5.52%	5.80%	5.28%
CVaR(0.05)	4.44%	4.73%	4.53%	4.44%	4.58%	**4.37%**	4.38%	**4.37%**	5.08%
Cvar(0.01)	6.80%	6.95%	6.82%	6.10%	7.07%	**6.83%**	6.45%	**6.77%**	7.11%
MaxDD	-7.48%	-8.83%	-7.54%	-6.90%	-8.57%	-9.57%	-9.40%	-9.39%	-10.01%
Cret	1.256	1.103	1.191	1.345	1.457	1.791	1.965	1.983	1.935
Cret_cost	0.845	0.557	0.740	0.889	0.888	1.598	1.803	1.801	1.923
Ircost	-0.054	-0.342	-0.152	-0.017	-0.013	0.417	**0.497**	**0.507**	0.487
Turnover	1.59	2.74	1.91	1.66	1.99	0.46	**0.35**	**0.39**	**0.02**

The best results for each statistical evaluation are labeled in bold. Turnover above 100% is possible because of short sales.

Table 3.3. Performance of V, LAD, Huber, Rnakcov, FAST-MCD, QCIQR, 2D-Winsorization, 2D-Huber models and Market index for $\mu_p = 20\%$

20%	**V**	**LAD**	**Huber**	**Rankcov**	**FAST-MCD**	**QCIQR**	**2D-winsor**	**2D-Huber**	**Market**
mean	0.060%	0.031%	0.057%	0.078%	0.089%	0.133%	0.155%	0.155%	0.160%
STD	1.978%	2.104%	1.999%	1.968%	2.035%	1.964%	2.007%	1.950%	2.343%
IR	0.217	0.105	0.205	0.286	0.315	0.489	0.556	**0.572**	0.491
VaR(0.05)	3.09%	3.59%	3.16%	3.22%	3.17%	3.17%	3.22%	3.06%	4.06%
Var(0.01)	6.05%	5.99%	6.17%	5.67%	6.20%	5.86%	5.57%	6.02%	5.28%
CVaR(0.05)	4.48%	4.98%	4.58%	4.49%	4.62%	4.39%	4.40%	4.39%	5.08%
Cvar(0.01)	6.84%	6.99%	6.85%	6.08%	7.09%	6.85%	6.46%	6.79%	7.11%
MaxDD	-7.35%	-9.23%	-7.33%	-6.51%	-8.51%	-9.29%	-9.15%	-9.09%	-10.01%
Cret	1.221	1.044	1.201	1.342	1.406	1.766	1.959	1.970	1.935
Cret_cost	0.816	0.524	0.721	0.877	0.851	1.555	1.777	1.770	1.923
Ircost	-0.077	-0.371	-0.165	-0.026	-0.042	0.396	0.486	0.493	0.487
Turnover	1.61	2.76	2.05	1.70	2.02	0.51	0.39	0.43	0.02

Table 3.4. GMVP performance of V, FAST-MCD, QCIQR, 2D-Winsorization, 2D-Huber models and Market index

GMVP	**V**	**Rankcov**	**FAST-MCD**	**QCIQR**	**2D-winsor**	**2D-Huber**	**Market**
mean	0.094%	0.084%	0.124%	0.149%	0.166%	**0.167%**	0.160%
STD	1.958%	1.943%	2.031%	2.037%	2.090%	2.032%	2.343%
IR	0.345	0.310	0.441	**0.526**	**0.574**	**0.592**	0.491
VaR(0.05)	3.05%	3.15%	3.04%	**3.28%**	3.35%	3.30%	4.06%
Var(0.01)	5.78%	5.09%	6.16%	**5.27%**	5.10%	**4.84%**	5.28%
CVaR(0.05)	**4.42%**	**4.35%**	4.53%	4.50%	4.45%	4.45%	5.08%
Cvar(0.01)	6.76%	6.40%	7.04%	6.93%	**6.42%**	**6.52%**	7.11%
MaxDD	-8.12%	-8.68%	-8.69%	-11.12%	-10.80%	-10.89%	-10.01%
Cret	1.451	1.382	1.677	1.894	**2.059**	**2.075**	1.935
Cret_cost	0.981	0.927	1.024	1.744	**1.948**	**1.939**	1.923
Ircost	0.056	0.013	0.090	0.468	**0.536**	**0.544**	0.487
Turnover	1.57	1.60	1.98	0.33	**0.22**	**0.27**	**0.02**

The replacement of squared error cost function with either absolute deviation or Huber functions yields negative results (LAD IR = 0.143, Huber IR = 0.199 vs. V IR = 0.239 at $\mu_p = 15\%$) compared with simple mean variance, which is consistent with the failure of some of the earlier LAD applications in asset allocations. Another traditional robust method Spearman Rank Correlation based covariance, though used by some practitioners, does little to improve the portfolio performance compared with simple mean and covariance matrix either. All these lackluster results of traditional robust measures contributed to the limited use of robust statistics in practice.

On the contrary, all four more recent robust estimations FAST-MCD, QCIQR, 2D-Winsorization and 2D-Huber yield solid improvements than simple mean and variance. All four methods have significantly higher mean returns (e.g. GVMP returns 0.094%, 0.124%, 0.149%, 0.166% and 0.167% for V, FAST-MCD, QCIQR, 2D-Winsorization and 2D-Huber respectively) with similar standard deviation (e.g. GVMP standard deviation1.958%, 2.031%, 2.037%, 2.09% and 2.032% for V, FAST-MCD, QCIQR, 2D-Winsorization and 2D-Huber respectively) and as a result higher information ratio.

The benefit of FAST-MCD is more modest compared with pair-wise robust estimations. The reason likely lies in the strict assumptions of the MCD approaches. Although both MCD methods and pair-wise robust estimators are designed to eliminate the effects of outliers, MCD models use a restrictive contamination model assuming complete dependence of outliers for different assets. In reality, although the returns of each industry sector are influenced by general market conditions, market risk only contributes to a small part of total variance of each asset. Pair-wise robust estimators offer more flexibility to calculate the covariance between two assets by looking at 2-demensional outliers instead of N-dimensional ones. Once the semi-positive definiteness property of the covariance matrix is guaranteed through transformation, they provide the best results among all the estimators for covariance matrix, including the models we discussed in Chapter 2. In every case we investigated, QCIQR, 2D-Winsorization and 2D-Huber yield results better than the market index. CCC-GARCH and DCC-GARCH (the best performing models in Chapter 2) only yield information ratios of 0.497 and 0.491 for the GMVP portfolio, which is less than QCIQR, 2D-Winsorization and 2D-Huber methods (IR = 0.526, 0.574 and 0.592 respectively). Compared with other estimation methods, QCIQR, 2D-Winsorization and 2D-Huber also give the lowest asset turnovers (0.33, 0.22 and 0.27 respectively), which indicates the stability of these models. In contrast, neither CCC-GARCH nor DCC-GARCH is a stable model. When the large turnover is taken into consideration for CCC-GARCH and DCC-GARCH (1.38 and 1.37), the transaction costs will further diminish the effectiveness of these models. As a result, the information ratios with transaction costs incorporated for QCIQR, 2D-Winsorization and 2D-Huber methods (0.468, 0.536 and 0.544 respectively) are significantly higher than CCC-GARCH and DCC-GARCH (0.196 for both). All these results clearly show the value of these new robust statistical methods.

Between QCIQR and 2D-Winsorization/2D-Huber, 2D-Winsorization/2D-Huber is believed to have more advantage over QCIQR. QCIQR is founded on a nonparametric approach, which incorporates less information compared with 2D-Huber. By shrinking the outliers toward the two-dimensional ellipsoid, 2D-Winsorization/2D-Huber reduces the unduly influence of outliers, yet has the ability to incorporate the parametrical information conveyed by these data points, which explains its best performance compared with other robust estimations. Between 2D-Winsorization and 2D-Huber, the 2D-Huber method indeed gives slightly better results than 2D-Winsorization, although it requires more computation time because of the iteration and requires the selection of a threshold value to stop the iteration. All these methods also can be implemented in a reasonable time-frame. The estimation of 500 covariance matrices took 15 min for QCIQR, 35 min for 2D-Winsorization, 3 hours for 2D-Huber and 10 hours for FAST-MCD. All these covariance matrices only need to be computed once and applied to different μ_p s.

3.7. Conclusion

In this Chapter, we investigated some of the traditional robust portfolio estimators: LAD, Huber and Rank correlation. The results showed that none of these methods provides significant improvement in portfolio performance. Then we implemented four of the recently developed robust covariance matrix estimators FAST-MCD, QCIQR, 2D-Winsorization and 2D-Huber, which to our knowledge have not been applied to asset allocation problems. All three methods yield significant improvement compared with simple mean-variance optimization. The pair-wise covariance matrix estimators, especially the 2D-Winsorization and 2D-Huber estimators, are able to outperform the market in the long run. These results indicate the potential value of modern robust statistics in asset allocation.

Chapter 4

Market-weight Bayesian Approaches

4.1. Introduction

As we have discussed in Chapter 2, when the number of observations T is of the same order of magnitude as the number of assets N, the estimates for the mean and covariance matrix are often highly unreliable and the 'optimal' asset weights based on these estimates are imprecise as well. Motivated from a Bayesian perspective, shrinkage methods combining the highly-structured market model and the sample mean/covariance try to reach the right balance between estimation error and bias. Nevertheless as shown in Chapter 2, when the number of days T is only 2-fold of the number of assets N, the method fails to achieve satisfactory results. In Chapter 3, we have tackled the problem using robust estimations by reducing or eliminating the effect of outliers in expected return and covariance matrix estimation. The results show that some of the most recent robust statistical methods can make significant contribution to improve portfolio performance. In this Chapter, we will further investigate alternative robust portfolio estimations by further exploring Bayesian approaches.

Shrinkage models combine the views of factor models/market equilibrium with sample mean/covariance of historical data to estimate expected returns and covariance matrices. In Chapter 2, we discussed some of the methods to shrink the sample covariance towards the covariance matrix estimated from market portfolio returns. Similar approaches can also be applied to the expected returns. The first Bayes-Stein estimation procedure was developed by Jorion (*42, 43*), in which the tangent portfolio is shrunk toward the global minimum-variance portfolio. The global minimum-variance portfolio does not depend on expected returns of the assets. Therefore it is not directly subject to the expected return estimation error from historical data. Jorion specified the prior that all assets have the same expected returns:

$$\Pi \sim \Phi(e\mu_0, \Sigma / \lambda) \tag{4.1}$$

where $\Phi(\mu, \sigma^2)$ represents a normal distribution; Π is a $N \times 1$ vector, the prior expected returns; e is $N \times 1$ vector of ones; μ_0 is the expected return of the global minimum-variance portfolio; Σ is $N \times N$ covariance matrix

λ determines the prior precision, which can be estimated from the data by using

$$\lambda = \frac{N+2}{(\bar{R} - e\mu_0)^T \Sigma^{-1} (\bar{R} - e\mu_0)} \tag{4.2}$$

The denominator measures the observed dispersion of the sample means around the common mean. Combining the prior with the sample means $R \sim \Phi(\bar{R}, \Sigma / T)$, the posterior expected returns can be expressed as:

$$E[R] = \frac{\lambda}{\lambda + T} e\mu_0 + \frac{T}{\lambda + T} \bar{R} \tag{4.3}$$

An obvious drawback is that the method imposes the assumption that all assets have the same expected return regardless of their risk profiles, which directly contradicts the common belief that higher systematic risks must be compensated with higher returns. Also the estimation of the global minimum-variance portfolio and its return is still based on historical returns. So a more economically sound approach was developed to incorporate asset pricing models in estimating prior expected returns. Asset pricing models such as the capital asset pricing model (CAPM) or arbitrage pricing theory (APT) with adjusted factor coefficients (e.g. risk factor coefficients estimated using market consensus instead of simple linear regression on historical data) are sometimes used to estimate expected returns. Pastor et al. (*44, 45*) applied expected returns calculated from asset pricing models as the prior. However, none of these models attracted much attention in practice, probably because of the poor performance of these estimators.

4.2. Black-Litterman Model

Originally introduced by Fisher Black and Robert Litterman in early 1990s (*46, 47*) at Goldman Sachs, the Black-Litterman model is another Bayes-Stein estimator (*42*) that combines the subjective views of investors regarding the expected returns of one or more assets with the market equilibrium expected returns (as the prior) to generate a mixed estimate of expected returns. As a proprietary trading model, none of the few available papers published by Goldman Sachs (*48-50*) discusses the model construction in detail, especially how to construct the views and estimate the views' covariance. So in this thesis, our interpretation is based on the publicly available information, Idzorek's guide (*51*) and our own understanding of the Bayesian statistics.

The Black-Litterman model revolutionizes the Bayesian shrinkage estimation of expected returns and covariance matrix by fully utilizing the market weights of each asset. The aforementioned shrinkage models all rely on market portfolio returns (or some other factors), which are calculated as market-capital-weighted average returns (or factor returns) of each individual asset. Instead of using the N assets' market weights directly, the weight information is aggregated into the final market return. It is our belief that such aggregation diminishes the information incorporated in the market weights and as a result contributes to the lackluster performance of the portfolio derived from these estimators.

Black-Litterman model estimates the expected return as a weighted average of the implied equilibrium return vector Π and the view vector Q, in which the relative weights depends on the relative confidence of the view versus the confidence in the implied equilibrium vector.

As discussed in Chapter 1, the quadratic utility function can be expressed as $U = w'\mu - \frac{1}{2}\lambda w'\Sigma w$. To maximize the utility:

$$\frac{\partial U}{\partial w} = \mu - \lambda \Sigma w = 0 \Rightarrow w = (\lambda \Sigma)^{-1} \mu . \tag{4.4}$$

If the market weights are used as equilibrium, through reverse optimization, the equilibrium return can are derived as:

$$\Pi = \lambda \Sigma w_{mkt} \tag{4.5}$$

Where Σ is the covariance matrix of the returns, an $N \times N$ matrix; w_{mkt} is the market capitalization weight, an $N \times 1$ vector; Π is the implied equilibrium return vector, an $N \times 1$ vector.

The market weights of different assets not only affect the market returns, but also influence the estimated market equilibrium returns of other assets indirectly through the covariance matrix. The expected return vector $E[R]$ based on market equilibrium can be expressed as a joint-normal distribution $E[R] \sim \Phi(\Pi, \tau\Sigma)$. The published Black-Litterman models still use the simple sample covariance matrix as the estimator, however more sophisticated models such as GARCH estimators may yield better results. A scalar τ is introduced to reduce the scale of covariance matrix since the uncertainty in the expected return should be much less volatile than individual returns for each period. Black and Litterman addressed the issue by setting the scalar τ close to zero. Nevertheless there are no mathematical arguments in Goldman's literature as to how to choose τ and the scalar τ is often determined by experience. The value of τ is typically set to between 0.01 and 0.05, and the model is calibrated based on the targeted level of tracking error. Intuitively, if one considers the expected return as the mean of historical returns, then τ can be estimated as $1/\sqrt{T}$ since the standard error of the mean is $1/\sqrt{T}$ of the standard deviation of the samples. If we are using an exponentially weighted moving average to estimate the covariance matrix, T can be adjusted to $T^* = \sum_{i=1}^{T} \lambda^{i-1}$ instead.

The views in the model are often expressed in a matrix format as

$$P \cdot E[R] = Q \tag{4.6}$$

Theoretically P can be any matrix with *N* columns. It is often used to express subjective views. For example, if the investor has a simple view that asset *i* will outperform asset *j* by x% during the next period, then equation is simply $[\cdots \quad P_i = 1 \quad \cdots \quad P_j = -1 \quad \cdots] \times E[R] = Q = x\%$.

More generally, there are usually prediction errors in investors' views. So the condition is no longer simply $P' E[R] = Q$, but $P' E[R] = Q + \varepsilon$, with the error term vector ε representing the estimation error of the views. If we assume these error terms follow a joint-normal distribution with mean 0 and covariance matrix Ω, then the view essentially follows a normal distribution

$$P' E[R] \sim \Phi(Q, \Omega). \tag{4.7}$$

Now from a Bayesian approach, the goal is to estimate a posterior expected return vector based on the prior distribution $E[R] \sim \Phi(\Pi, \tau\Sigma)$ and the condition $P' E[R] \sim \Phi(Q, \Omega)$. To yield the new combined expected return vector, the original problem can be formulated as a generalized least square regression question (unlike simple linear square regression that assumes all residuals follow the same distribution $\Phi(0, \sigma^2)$, generalized least square regression allows the residuals to have different variances.):

$$\begin{pmatrix}\Pi\\Q\end{pmatrix}=\begin{pmatrix}I\\P\end{pmatrix}E[R]+\varepsilon \tag{4.8}$$

where ε has mean 0 and covariance matrix $W=\begin{bmatrix}\tau\Sigma & 0\\ 0 & \Omega\end{bmatrix}$.

let $Y=\begin{pmatrix}\Pi\\Q\end{pmatrix}$, $X=\begin{pmatrix}I\\P\end{pmatrix}$ and treat $E[R]$ as β, we can apply the generalized linear least square model: $\hat{\beta}=(X'W^{-1}X)^{-1}X'W^{-1}Y$ and yield the following expected value for $E[R_\Omega]$:

$$E[R_\Omega]=\left[\begin{pmatrix}I & P'\end{pmatrix}\begin{pmatrix}(\tau\Sigma)^{-1} & 0\\ 0 & \Omega^{-1}\end{pmatrix}\begin{pmatrix}I\\P\end{pmatrix}\right]^{-1}\left[\begin{pmatrix}I & P'\end{pmatrix}\begin{pmatrix}(\tau\Sigma)^{-1} & 0\\ 0 & \Omega^{-1}\end{pmatrix}\begin{pmatrix}\Pi\\Q\end{pmatrix}\right]\Rightarrow$$

$$E[R_\Omega]=\left[(\tau\Sigma)^{-1}+P'\Omega^{-1}P\right]^{-1}\left[(\tau\Sigma)^{-1}\Pi+P'\Omega^{-1}Q\right] \tag{4.9}$$

$$Var(E[R_\Omega])=(X'W^{-1}X)^{-1}=\left[(\tau\Sigma)^{-1}+P'\Omega^{-1}P\right]^{-1}$$

and the corresponding new weight allocation vector can be estimated as

$$w=(\lambda\Sigma)^{-1}E[R_\Omega]. \tag{4.10}$$

Determining the covariance matrix for the error term of the views is the most abstract and complicated aspect of the Black-Litterman model. There is no consensus as to how to estimate the covariance matrix Ω. On the one hand, it increases the flexibility of model since quantitative investors can express slightly different views through mathematical formation. On the other hand, such flexibility increases the difficulty of applying Black-Litterman model since the investors have to specify the multivariate normal distribution on different views. Interestingly, in most of the Black-Litterman model papers (actually in all papers we have identified), Ω is set to a diagonal covariance matrix assuming that the views are independent of each other. It is our belief that such an assumption is overly restrictive. Even for subjective views, many of them are not independent since the views can share some of the same contributing factors. The addition of off-diagonal elements adds further flexibility to the model. For example, two investors can agree on the same two views and their variance, but differ in their correlation. The higher correlation of the two views can be expressed as a higher covariance term in Ω.

To extend the Bayesian idea behind the Black-Litterman model to pure quantitative models, one can again treat the asset returns from the market weights as the prior. Instead of subjective views, we can express the historical sample means (or more robust estimates) as the views and the final conditional results can be estimated from the prior and views. This approach, although intuitively appealing, is difficult to implement since there are many parameters, namely λ, Σ, τ and Ω to choose. There are few guidelines behind choosing these parameters and practitioners often rely on empirical results to select them, which makes the models difficult to implement and subject to data snooping.

4.3. Regularization Models

A similar approach that fully incorporates the market weight information is regularization methods, which eliminate the multiple parameters that we have to choose for the Black-Litterman model. Lauprete showed that market weights could be used to increase the portfolio performance through regularization (*29*). In general, the family of regularized portfolio estimators can be expressed as:

$$(w(\lambda), q(\lambda)) = \arg \min_{Aw=b, q\in R} \left(\frac{1}{T} \sum_{t=1}^{T} \eta(w' R_t - q) + \lambda \|w - w_m\|_p^p \right) \tag{4.11}$$

where $\lambda \geq 0$ is the regularization parameter; $\|\cdot\|_p$ is the $L_p - norm$ in R^N for $w \in R^N$; w_m is the market weight, which will serve as a prior portfolio in the regularization approach.

Normally, only p = 1 and 2 are considered, which correspond respectively to L_1 and L_2 regularization. The intuition is clear in the new cost functions that we try to minimize. In Chapter 3, we examined three cost functions: variance, LAD and Huber function. In this section, we add the market portfolio as a prior. If the market is efficient, we should penalize the final cost function if the proposed asset weights deviate from the prior. As a result, extreme deviations from the prior are unlikely. Unlike the Black-Litterman model, the only parameter needed for regularization method is the penalty coefficient λ. The term $\lambda \|w - w_m\|_p^p$ reflects the investor's a priori confidence in portfolio w_m. A large λ means large penalty for any deviation and strong confidence in w_m; a small λ reflects weak confidence in w_m. An appropriate choice of λ will reduce the variability of the portfolio estimate, without biasing the estimate too much.

In this thesis, we choose the parameter λ using 5-fold cross validation. For any given λ, we implement the following steps

1. Divide the T observations into 5 subsets of $T/5$ observations. Call these subsets $T(i)$ for $i = 1, \cdots, 5$.
2. For every $i = 1, \cdots, 5$, run the optimization to yield the optimal $(\hat{w}(\lambda), \hat{q}(\lambda))$ for the in-sample data: $(\hat{w}(\lambda), \hat{q}(\lambda)) = \arg \min_{Aw=b, q\in R} \left(\frac{1}{0.8T} \sum_{t \in T \setminus T(i)} \eta(w' R_t - q) + \lambda \|w - w_m\|_p^p \right)$.

3. For every $i = 1, \cdots, 5$ apply the $(\hat{w}(\lambda), \hat{q}(\lambda))$ to the out-of-sample data to calculate a sum of squared errors $PE_\lambda(i) = \sum_{t \in T(i)} [\hat{w}_\lambda(i)' R_t - \hat{q}_\lambda(i))^2]$.

4. Calculate the total sum of squared errors $PE_\lambda = \sum_{i=1}^{5} PE_\lambda(i)$.

A series of candidate values of λ from 0.01 to 2 are tested to yield a value of λ with minimum total sum of squared errors PE_λ.

In Chapter 3, we discussed three functions $\eta(x)$ for cost function $\frac{1}{T}\sum_{t=1}^{T}\eta(w' R_t - q)$, namely the square function, absolute value and Huber loss function. We can combine each function with the penalty term $\lambda \|w - w_m\|_p^p$ using both L_1 ($p = 1$) and L_2 ($p = 2$) regularization.

4.4. Application of Regularization Models

Since the regularization methods do not rely on the subjective choices of many parameters as Black-Litterman model does, we will focus on regularization methods and study the performance of the following six estimators:

- V1: L_1-regularized variance;
- V2: L_2-regularized variance;
- LAD1: L_1-regularized least-absolute deviations;
- LAD2: L_2-regularized least-absolute deviations;
- H1: L_1-regularized Huber estimator (α is set to 0.10);

- H2: L_2-regularized Huber estimator (α is set to 0.10);

Again, the application uses the same daily return data on 51 MSCI industry sector indexes, from 01/03/1995 to 02/07/2005.

The following are the model implementation details:

V1 (L_1-regularized Variance)

The original formulation

$$\begin{aligned}&\min_{w}\ w'\Sigma w+\lambda\left\|w-w_m\right\|_1^1\\&s.t.\ w'\bar{R}\geq\mu_p,\ w'e=1\end{aligned}\tag{4.12}$$

can be transformed to a quadratic programming problem

$$\begin{aligned}&\min_{w,q,z,y}\ w'\Sigma w+\lambda\sum_{j=1}^{N}y_j\\&s.t.\quad w'e=1\\&\qquad w'\bar{R}\geq\mu_p && \Rightarrow\quad -\bar{R}'w\leq-\mu_p\\&\qquad y_j\geq w_j-w_{j,m},\ j=1,\cdots,N && \Rightarrow\quad w_j-y_j\leq w_{j,m}\\&\qquad y_j\geq -w_j+w_{j,m},\ j=1,\cdots,N && \Rightarrow\quad -w_j-y_j\leq -w_{j,m}\end{aligned}\tag{4.13}$$

that has $2N$ variables and $2+2N$ constraints.

V2 (L_2-regularized Variance)

The original V2 model is expressed as a QP problem:

$$\min_{w} w'\Sigma w+\lambda\sum_{j=1}^{N}(w_j-w_{m,j})^2 \quad \Leftrightarrow \min_{w} \tfrac{1}{2}w'(\Sigma+\lambda I)w-(\lambda w_m)'w$$
$$s.t.\ w'\overline{R}\ge\mu_p,\ w'e=1 \tag{4.14}$$

LAD1 (L_1-regularized LAD)

LAD1 is essentially L_1-regularized α-Shortfall with $\alpha = 0.5$

$$\min_{w,q} \tfrac{1}{T}\sum_{t=1}^{T}(w'R_t-q)-\tfrac{1}{T}\sum_{t=1}^{T}\tfrac{1}{\alpha}(w'R_t-q)I_{\{w'R_t\le q\}}+\lambda\|w-w_m\|_1^1$$
$$s.t.\quad w'\overline{R}\ge\mu_p,\ w'e=1 \tag{4.15}$$

L_1-regularized α-Shortfall can be converted to a LP Formulation:

$$\begin{aligned}
&\min_{w,q,z,y}\ w'\overline{R}-q+\tfrac{1}{\alpha T}\sum_{t=1}^{T}z_i+\lambda\sum_{j=1}^{N}y_j\\
&s.t.\quad w'\overline{R}\ge\mu_p,\ w'e=1\\
&\qquad z_t\ge q-w'R_t,\ t=1,\cdots,T\\
&\qquad y_j\ge w_j-w_{j,m},\ j=1,\cdots,N\\
&\qquad y_j\ge -w_j+w_{j,m},\ j=1,\cdots,N\\
&\qquad z_t\ge 0,\ i=1,\cdots,T
\end{aligned} \tag{4.16}$$

that has $2N+1+T$ variables and $M+2T+2N$ constraints, with $N+1$ dense columns.

LAD2 (L_2-regularized LAD)

LAD2 is essentially L_2-regularized α-Shortfall with $\alpha = 0.5$

$$\min_{w,q} \tfrac{1}{T}\sum_{t=1}^{T}(w'R_i-q)-\tfrac{1}{T}\sum_{t=1}^{T}\tfrac{1}{\alpha}(w'R_t-q)I_{\{w'R_t\le q\}}+\lambda\|w-w_m\|_2^2$$
$$s.t.\quad w'\overline{R}\ge\mu_p,\ w'e=1 \tag{4.17}$$

QP Formulation for L_2-regularized α-Shortfall

$$
\begin{aligned}
&\min_{w,q,z} \; w'\bar{R} - q + \tfrac{1}{\alpha T}\sum_{t=1}^{T} z_t + \lambda \sum_{j=1}^{N}(w_j - w_{m,j})^2 \\
&s.t. \quad w'e = 1 \\
&\qquad w'\bar{R} \ge \mu_p && \Rightarrow \quad -\bar{R}'w \le -\mu_p \\
&\qquad z_t \ge q - w'R_t,\; t = 1,\cdots,T && \Rightarrow \quad -R_t'w + q - z_t \le 0 \\
&\qquad z_t \ge 0,\; t = 1,\cdots,T && \Rightarrow \quad -z_t \le 0
\end{aligned}
\tag{4.18}
$$

that has $N+1+T$ variables and $M+2T$ constraints, with $N+1$ dense columns.

H1 (L_1-regularized Huber Portfolio Estimator)

We can construct H1 as a quadratic programming problem:

$$
\begin{aligned}
&\min_{w,q,t} \; \tfrac{1}{T}\sum_{t=1}^{T} u_t^2 + \tfrac{1}{T}\sum_{t=1}^{T} 2\gamma(z_t^+ + z_t^-) + \lambda \sum_{j=1}^{N} y_j \\
&s.t. \quad w'e = 1 \\
&\qquad w'\bar{R} \ge \mu_p && \Rightarrow \quad -\bar{R}'w \le -\mu_p \\
&\qquad z_t^+ - z_t^- = w'R_t - q - u_t,\; t = 1,\cdots,T && \Rightarrow \quad w'R_t - q - u_t - z_t^+ + z_t^- = 0 \\
&\qquad y_j \ge w_j - w_{j,m},\; j = 1,\cdots,N \\
&\qquad y_j \ge -w_j + w_{j,m},\; j = 1,\cdots,N \\
&\qquad z^+, z^- \ge 0
\end{aligned}
\tag{4.19}
$$

H2 (L_2-regularized Huber Portfolio Estimator)

Similarly we can formulate H2 as a quadratic programming problem:

$$
\begin{aligned}
&\min_{w,q,t} \tfrac{1}{T}\sum_{t=1}^{T} u_t^2 + \tfrac{1}{T}\sum_{t=1}^{T} 2\gamma(z_t^+ + z_t^-) + \lambda\sum_{j=1}^{N}(w_j - w_{m,j})^2 \Rightarrow \\
&\min_{w,q,t} \tfrac{1}{T}\sum_{t=1}^{T} u_t^2 + w'(\lambda D)w + \tfrac{1}{T}\sum_{t=1}^{T} 2\gamma(z_t^+ + z_t^-) - 2\lambda w_m' w \\
&s.t.\quad w'e = 1 \\
&\qquad w'\overline{R} \geq \mu_p \qquad\qquad\qquad\quad \Rightarrow \quad -\overline{R}'w \leq -\mu_p \\
&\qquad z_t^+ - z_t^- = w'R_t - q - u_t,\ i = 1,\cdots,T \ \Rightarrow \quad w'R_t - q - u_t - z_t^+ + z_t^- = 0 \\
&\qquad z^+, z^- \geq 0
\end{aligned}
\tag{4.20}
$$

that has $N+1+3T$ variables and $M+3T$ constraints, with $N+1$ dense columns.

Again a range of target expected portfolio returns from 10% to 50% annual rate are used for portfolio construction. Appendix 4 shows detailed return statistics for V1, V2, LAD1, LAD2, H1 and H2. Table 4.1 – 4.2 show the summarized results for annual expected return 15% and 20%.

Table 4.1. Performance of V, V1, V2, LAD1, LAD2, H1, H2 models and Market index for $\mu_p = 15\%$.

15%	**V**	**V1**	**V2**	**LAD1**	**LAD2**	**H1**	**H2**	**Market**
mean	0.065%	0.198%	0.140%	0.199%	0.137%	0.191%	0.147%	0.160%
STD	1.962%	2.431%	2.021%	2.430%	1.924%	2.388%	2.158%	2.343%
IR	0.239	**0.589**	0.498	**0.591**	0.513	0.576	0.490	0.491
VaR(0.05)	**3.06%**	3.71%	3.24%	3.71%	**3.21%**	3.66%	3.53%	4.06%
Var(0.01)	5.78%	6.65%	5.88%	6.65%	**5.82%**	6.66%	5.91%	**5.28%**
CVaR(0.05)	4.44%	5.05%	4.47%	5.04%	**4.30%**	5.00%	4.69%	5.08%
Cvar(0.01)	6.80%	7.31%	**6.56%**	7.28%	**6.34%**	7.27%	6.76%	7.11%
MaxDD	**-7.48%**	-8.35%	-8.09%	-8.35%	**-7.70%**	-8.35%	-8.40%	-10.01%
Cret	1.256	**2.328**	1.814	**2.338**	1.807	2.250	1.853	1.935
Cret_cost	0.845	**2.252**	1.756	**2.262**	1.736	2.174	1.790	1.923
Ircost	-0.054	**0.569**	0.475	**0.572**	0.484	0.555	0.467	0.487
Turnover	1.59	**0.13**	**0.13**	**0.13**	0.16	0.14	0.14	**0.02**

The best results for each statistical evaluation are labeled in bold. Turnover above 100% is possible because of short sales.

Table 4.2. Performance of V, V1, V2, LAD1, LAD2, H1, H2 models and Market index for $\mu_p = 20\%$.

20%	V	V1	V2	LAD1	LAD2	H1	H2	Market
mean	0.060%	**0.207%**	**0.142%**	**0.205%**	0.139%	0.198%	0.149%	0.160%
STD	1.978%	2.465%	2.023%	2.464%	1.927%	2.423%	2.165%	2.343%
IR	0.217	**0.606**	0.508	**0.601**	0.521	0.590	0.496	0.491
VaR(0.05)	**3.09%**	3.67%	3.22%	3.67%	**3.20%**	3.61%	3.40%	4.06%
Var(0.01)	6.05%	6.54%	**5.97%**	6.54%	**5.87%**	6.54%	6.03%	5.28%
CVaR(0.05)	4.48%	5.09%	**4.48%**	5.08%	**4.32%**	5.04%	4.72%	5.08%
Cvar(0.01)	6.84%	7.36%	**6.57%**	7.34%	**6.35%**	7.27%	6.90%	7.11%
MaxDD	**-7.35%**	-8.15%	-7.90%	-8.15%	**-7.54%**	-8.16%	-8.20%	-10.01%
Cret	1.221	**2.421**	1.839	**2.399**	1.826	2.329	1.872	1.935
Cret_cost	0.816	**2.323**	1.768	**2.303**	1.744	2.231	1.796	1.923
Ircost	-0.077	**0.582**	0.480	**0.577**	0.486	0.565	0.468	0.487
Turnover	1.61	**0.17**	0.16	**0.16**	0.18	0.17	0.17	**0.02**

Overall, the regularization methods are computationally intensive methods compared with their counterparts V, LAD and Huber. The addition of the penalty term extends the dimension of the optimization problems and increases the number of constraints. The cross-validation of each penalty coefficient λ increases the computation further by ~25 fold. Unlike robust estimation of mean and covariance matrix, which only need to calculate the parameters once for all μ_p, the optimization problem needs to be performed for every μ_p. As a result, the running time for each regularization application on a fixed μ_p is long ranging from ~3 hours for LAD1 to 7 days for H1.

These regularization methods, though computationally intensive, carries great advantages. All regularization methods at least double the information ratio compared with the simple mean-variance approach. The L_2 regularization methods usually offer slightly better IR than market index, while the L_1 regularization methods provide more solid increase (IR = 0.589, 0.591, 0.576 and 0.491 for V1, LAD1, H1 and Market index respectively at $\mu_p = 15\%$), which is comparable to the result of 2D-Winsorization and 2D-Huber. However, the regularization further decreases the asset turnover (turnover = 0.17, 0.16, 0.17, 0.34, 0.39 and 1.59 for V1, LAD1, H1, 2D-Winsorization, 2D-Huber and V respectively at $\mu_p = 15\%$), which indicates further decreases in the trading costs and the stability of the models. Among all estimation and optimization methods, L_1 methods offer the highest information ratios and lowest turnovers. Overall L_1 regularization methods outperform the robust covariance estimation methods, such as QCIQR and 2D-Huber methods. However they achieve so at the cost of significantly higher computational complexity compared with QCIQR and 2D-Huber methods.

4.5. Conclusion

If the market index is used as the prior, our results showed that regularization methods, especially L_1 regularization methods improve the performance of portfolio estimators. When the number of assets is of then same order of magnitude as the number of observations, the profit of using regularization methods is significant.

Chapter 5

Summary and Future Research Directions

Mean-variance portfolio optimization is the foundation of modern portfolio theory and is the most cited method in asset allocation literature. However, the implementation of the mean-variance efficient frontier is limited in practice by difficulties in estimating model inputs, expected returns and the covariance matrices of different assets, and the sensitivity of asset weights assigned to these inputs.

Traditionally, sample means and covariance matrices from historical data were used, which are subject to large estimation errors. This thesis investigates a variety of more sophisticated methods to address the estimation error problem. We first surveyed some of the traditional factor models and Bayesian shrinkage models, combining factor models with sample covariance to estimate covariance matrices. Our results show that these models have limited success because of the influence of outliers on factor models, especially when the number of data points is of the same order as the number of assets. To reduce or eliminate the effect of outliers, we then investigated some of robust statistical approaches with a focus on recently developed methods such as FAST-MCD, quadrant-correlation-based covariance and 2D-Huber-based covariance. These new robust methods, especially pair-wised robust covariance estimation models, are shown to be valuable tools in improving risk-adjusted portfolio performance and reducing asset turnover. Finally we investigated more complicated Bayesian models using market weights of each asset as priors, and implement regularization methods based on the Bayesian approach. These penalized robust methods also significantly increase the portfolio performance. Overall L_1

regularization methods outperform the robust covariance estimation methods, such as QCIQR, 2D-Winsorization and 2D-Huber methods. However, they achieve so at the cost of significantly higher computational complexity compared with QCIQR, 2D-Winsorization and 2D-Huber methods. In conclusion, robust asset allocation methods have great potential to improve risk-adjusted portfolio returns and therefore deserve further exploration in investment management research.

We submit that the methods covered in this thesis are far from complete. There are many robust statistical methods and combinations of different methods that are worthy of further investigation. For example, it is possible to combine pair-wise robust estimation of covariance matrix (*e.g.*, 2D-Huber) with regularization V1 or V2 by replacing the simple sample covariance matrix with the robust covariance matrix.

Considering the relatively high computational complexity of the regularization methods we have outlined, especially if we want to apply these methods to individual stock selection, future study may also focus on developing fast algorithms to efficiently solve these problems.

Appendix

Appendix 1. MSCI US Sector Indices

The MSCI US Equity Indices include sector, industry group and industry indices as classified in accordance with the Global Industry Classification Standard (GICS®) for each of the market capitalization indices described above. MSCI calculates hundreds of industry, industry group and sector indices for the MSCI US Equity Indices.

Final 51 industries:

Energy Equipment; Oil and Gas; Chemicals; Const. materials; Containers Metals and Mining; Paper Products; Aerospace; Building Products; Construction; Elec. Equipment; Industrial Machinery; Trading Companies; Comm. Services; Air Freight; Airlines; Road and Rail; Auto Components; Automobiles; Household Durables; Leisure Equipment; Textiles; Hotel; Media; Multiline Retail; Specialty Retail; Food and Staples; Beverages; Food Products; Tobacco House Products; Personal Products; Health Equipment; Health Providers; Pharmaceuticals; Div. Financials; Insurance; Real Estate; IT Services; Software Comm. Equipment; Computers; Electronics; Office Electronics; Semi Equipment; Div. Telecomm; Elec. Utilities; Gas Utilities; Multi-Utilities

Appendix 2. Performance of V, CAPM, Principal, Mahalanobis, Frobenius, CCC-GARCH, DCC-GARCH models

V

V	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.0682%	0.0649%	0.0595%	0.0548%	0.0509%	0.0484%	0.0465%	0.0427%
STD	1.9489%	1.9615%	1.9780%	2.0028%	2.0352%	2.0757%	2.1233%	2.2370%
IR	0.25238	0.23851	0.21703	0.19719	0.18029	0.16801	0.15804	0.13766
VaR(0.05)	3.0086%	3.0555%	3.0902%	0.031249	0.032479	0.033955	0.033618	0.03735
Var(0.01)	5.7800%	5.7800%	6.0548%	0.065358	0.06502	0.06288	0.060741	0.056463
CVaR(0.05)	4.4213%	4.4444%	4.4838%	0.045373	0.046177	0.047324	0.048478	0.051182
Cvar(0.01)	6.7802%	6.8034%	6.8409%	0.069196	0.069984	0.070771	0.071558	0.073133
MaxDD	-7.6079%	-7.4777%	-7.3476%	-0.075745	-0.078444	-0.081143	-0.083843	-0.089411
Cret	1.2787	1.256	1.2209	1.1892	1.1625	1.1431	1.1271	1.0919
Cret_cost	0.86406	0.84504	0.81609	0.78806	0.76159	0.73889	0.71719	0.66926
Ircost	-0.03826	-0.053567	-0.077492	-0.10004	-0.12043	-0.13632	-0.15021	-0.17938
Turnover	1.5702	1.5879	1.614	1.6487	1.6946	1.7485	1.8113	1.9617

CAPM

CAPM	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.0285%	0.0284%	0.0282%	0.0275%	0.0254%	0.0220%	0.0203%	0.0191%
STD	1.7498%	1.7542%	1.7612%	1.7715%	1.7868%	1.8096%	1.8406%	1.9234%
IR	0.11764	0.1166	0.11526	0.11187	0.10238	0.087656	0.079424	0.07174
VaR(0.05)	2.8935%	2.8040%	2.9298%	0.028231	0.028229	0.02928	0.029709	0.030648
Var(0.01)	5.0127%	5.0127%	5.0127%	0.050127	0.050127	0.050725	0.051413	0.052712
CVaR(0.05)	4.0866%	4.1110%	4.1514%	0.04193	0.042387	0.042911	0.043543	0.045245
Cvar(0.01)	6.2605%	6.2702%	6.2799%	0.062604	0.062205	0.061934	0.062037	0.062371
MaxDD	-7.5325%	-7.5325%	-7.5325%	-0.075551	-0.076035	-0.076518	-0.077002	-0.077968
Cret	1.0682	1.0668	1.0649	1.0604	1.0478	1.0282	1.0165	1.0029
Cret_cost	0.98884	0.98215	0.97393	0.96213	0.94234	0.91549	0.89522	0.86228
Ircost	0.054056	0.048668	0.042105	0.032685	0.016725	-0.0049038	-0.020159	-0.041594
Turnover	1.5702	1.5879	1.614	1.6487	1.6946	1.7485	1.8113	1.9617

Principal

Principal	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.0616%	0.0558%	0.0511%	0.0455%	0.0409%	0.0359%	0.0312%	0.0226%
STD	1.6722%	1.6756%	1.6864%	1.7053%	1.7311%	1.7636%	1.8018%	1.8944%
IR	0.2655	0.24	0.21863	0.19254	0.17049	0.14695	0.12504	0.08586
VaR(0.05)	2.8001%	2.7439%	2.7821%	0.027898	0.028034	0.02863	0.028988	0.030084
Var(0.01)	4.8291%	5.0248%	5.1371%	0.051413	0.051752	0.052092	0.052432	0.053111
CVaR(0.05)	3.7596%	3.7818%	3.8158%	0.038595	0.039202	0.039947	0.040721	0.04254
Cvar(0.01)	5.5555%	5.5968%	5.6649%	0.057788	0.058987	0.060186	0.061385	0.063782
MaxDD	-6.5682%	-6.5637%	-6.3892%	-0.066823	-0.071622	-0.076422	-0.081221	-0.090819
Cret	1.2684	1.2318	1.2025	1.1674	1.1383	1.1071	1.0777	1.0231
Cret_cost	1.0572	1.0221	0.99199	0.95619	0.92405	0.88948	0.85573	0.79065
Ircost	0.10836	0.079349	0.054059	0.023742	-0.0032451	-0.032035	-0.059586	-0.11036
Turnover	0.7288	0.7466	0.76976	0.79836	0.83413	0.87541	0.92256	1.0307

Mahalanobis

Mahalanobis	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.0466%	0.0410%	0.0366%	0.0320%	0.0267%	0.0224%	0.0184%	0.0103%
STD	1.5872%	1.5902%	1.5955%	1.6049%	1.6218%	1.6493%	1.6831%	1.7674%
IR	0.21157	0.18589	0.16526	0.14379	0.11885	0.097787	0.078998	0.041911
VaR(0.05)	2.7573%	2.7573%	2.7295%	0.027283	0.027727	0.028692	0.029128	0.030662
Var(0.01)	4.9286%	4.9580%	4.9580%	0.04961	0.049917	0.050224	0.050531	0.051145
CVaR(0.05)	3.8743%	3.8930%	3.9030%	0.039321	0.039741	0.040351	0.041134	0.04285
Cvar(0.01)	6.0227%	6.0599%	6.0546%	0.060478	0.060675	0.061651	0.062627	0.06458
MaxDD	-7.8854%	-7.8854%	-7.6723%	-0.074494	-0.072265	-0.070037	-0.073072	-0.082956
Cret	1.1847	1.1519	1.1262	1.1	1.0699	1.0444	1.0212	0.97324
Cret_cost	1.0519	1.0183	0.98981	0.95966	0.92508	0.89384	0.86385	0.80191
Ircost	0.10353	0.074105	0.048535	0.02109	-0.010599	-0.038525	-0.06461	-0.11636
Turnover	0.47586	0.49321	0.51671	0.54622	0.58222	0.62334	0.67	0.77505

Frobenius

Frobenius	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.0777%	0.0735%	0.0689%	0.0644%	0.0584%	0.0533%	0.0491%	0.0414%
STD	1.6694%	1.6732%	1.6807%	1.6925%	1.7138%	1.7421%	1.7776%	1.8657%
IR	0.33574	0.31684	0.29567	0.27447	0.24572	0.22051	0.19899	0.15986
VaR(0.05)	2.6663%	2.6894%	2.7249%	0.027458	0.027764	0.028287	0.029807	0.03205
Var(0.01)	5.1378%	5.1378%	5.1299%	0.051132	0.052349	0.054018	0.052603	0.054833
CVaR(0.05)	3.8408%	3.8801%	3.9215%	0.039679	0.040407	0.041241	0.042177	0.044196
Cvar(0.01)	5.9780%	5.9497%	5.8956%	0.058239	0.058714	0.059293	0.059872	0.062042
MaxDD	-7.2623%	-7.2623%	-7.1404%	-0.068876	-0.066348	-0.066741	-0.071812	-0.081954
Cret	1.3753	1.3462	1.3148	1.2843	1.244	1.2095	1.1806	1.1269
Cret_cost	1.1766	1.1472	1.1144	1.0811	1.0381	0.99917	0.96408	0.8957
Ircost	0.20097	0.17899	0.15381	0.12765	0.09342	0.062219	0.034458	-0.017936
Turnover	0.62518	0.64083	0.6623	0.69022	0.72475	0.76539	0.81155	0.91978

CCC-GARCH

CCC-GARCH	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.1067%	0.1026%	0.0989%	0.0957%	0.0918%	0.0891%	0.0874%	0.0816%
STD	1.6061%	1.6116%	1.6196%	1.6320%	1.6509%	1.6767%	1.7083%	1.7944%
IR	0.47915	0.45883	0.44047	0.42305	0.40099	0.38305	0.36875	0.32783
VaR(0.05)	2.5497%	2.5796%	2.6203%	0.025878	0.026342	0.027604	0.028403	0.029907
Var(0.01)	4.2254%	4.4185%	4.6115%	0.048046	0.049977	0.049911	0.050129	0.050173
CVaR(0.05)	3.6164%	3.6384%	3.6593%	0.036846	0.037325	0.038004	0.038723	0.040783
Cvar(0.01)	5.4970%	5.5060%	5.4998%	0.054936	0.054874	0.054811	0.055169	0.058408
MaxDD	-6.1495%	-6.1495%	-6.1495%	-0.061495	-0.061495	-0.061495	-0.061495	-0.070905
Cret	1.5981	1.5644	1.5354	1.5097	1.478	1.4548	1.4386	1.3871
Cret_cost	1.1302	1.1042	1.0809	1.0586	1.0313	1.0092	0.99123	0.94013
Ircost	0.16801	0.14695	0.12777	0.10928	0.086602	0.068454	0.054289	0.01531
Turnover	1.3862	1.3942	1.4046	1.4202	1.4394	1.4628	1.4897	1.5552

DCC-GARCH

DCC-GARCH	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.1059%	0.1015%	0.0977%	0.0945%	0.0905%	0.0877%	0.0860%	0.0802%
STD	1.6059%	1.6113%	1.6192%	1.6315%	1.6505%	1.6763%	1.7080%	1.7946%
IR	0.47566	0.45424	0.43519	0.41761	0.39552	0.3774	0.36323	0.32225
VaR(0.05)	2.5905%	2.6568%	2.6543%	0.026612	0.026461	0.027456	0.028393	0.030092
Var(0.01)	4.2790%	4.4723%	4.6656%	0.048588	0.050521	0.05008	0.049501	0.050086
CVaR(0.05)	3.6312%	3.6517%	3.6692%	0.036918	0.037365	0.038016	0.038742	0.040817
Cvar(0.01)	5.5055%	5.5155%	5.5099%	0.055043	0.054988	0.054932	0.05513	0.058418
MaxDD	-6.1470%	-6.1470%	-6.1470%	-0.06147	-0.06147	-0.06147	-0.06147	-0.071142
Cret	1.5918	1.5563	1.5262	1.5002	1.4686	1.4451	1.4291	1.3776
Cret_cost	1.1272	1.0998	1.0756	1.0531	1.0259	1.0036	0.98568	0.93449
Ircost	0.16557	0.14335	0.12343	0.10472	0.082002	0.06363	0.049547	0.010504
Turnover	1.3814	1.3895	1.4	1.4158	1.4351	1.4586	1.4858	1.5516

Appendix 3. Performance of LAD, Huber, Rnakcov, FAST-MCD, QCIQR, and 2D-Huber models

LAD

LAD	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.0390%	0.0405%	0.0308%	0.0532%	0.0573%	0.0468%	0.0363%	0.0336%
STD	2.0249%	2.0450%	2.1036%	2.1458%	2.2087%	2.2861%	2.3527%	2.4726%
IR	0.13896	0.14293	0.10546	0.1787	0.18715	0.14767	0.11126	0.097997
VaR(0.05)	3.4288%	3.4770%	3.5935%	0.035212	0.036235	0.035802	0.037372	0.040158
Var(0.01)	5.2196%	5.2205%	5.9874%	0.059573	0.062203	0.067012	0.071894	0.077593
CVaR(0.05)	4.6188%	4.7336%	4.9822%	0.05039	0.051664	0.054096	0.056243	0.058867
Cvar(0.01)	7.0873%	6.9545%	6.9893%	0.067981	0.069383	0.076859	0.082969	0.088189
MaxDD	-8.8324%	-8.8325%	-9.2305%	-0.081331	-0.078209	-0.083512	-0.093256	-0.1051
Cret	1.0966	1.1027	1.0437	1.1623	1.1785	1.1083	1.0432	1.0143
Cret_cost	0.55788	0.5572	0.52372	0.57431	0.57355	0.52602	0.48472	0.44834
Ircost	-0.34586	-0.34214	-0.37074	-0.29772	-0.28555	-0.32506	-0.36127	-0.3811
Turnover	2.7077	2.7352	2.7623	2.824	2.8847	2.9848	3.0693	3.2692

Huber

Huber	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.0657%	0.0545%	0.0567%	0.0315%	0.0422%	0.0348%	0.0191%	0.0390%
STD	1.9827%	1.9699%	1.9990%	2.0757%	2.0303%	2.0719%	2.1505%	2.4310%
IR	0.23878	0.19945	0.20456	0.10929	0.14983	0.12098	0.06397	0.11559
VaR(0.05)	3.0557%	3.0849%	3.1556%	0.033031	0.033865	0.034515	0.035135	0.038112
Var(0.01)	6.3967%	5.8765%	6.1745%	0.064663	0.062374	0.064247	0.068386	0.066634
CVaR(0.05)	4.5734%	4.5316%	4.5838%	0.047781	0.046069	0.047228	0.050533	0.055448
Cvar(0.01)	6.9638%	6.8172%	6.8466%	0.072177	0.068344	0.069482	0.074589	0.080385
MaxDD	-7.7900%	-7.5382%	-7.3316%	-0.08108	-0.074815	-0.077326	-0.083285	-0.099892
Cret	1.2581	1.1913	1.2012	1.0506	1.1136	1.0684	0.97941	1.0488
Cret_cost	0.70827	0.73952	0.72087	0.62316	0.69584	0.66663	0.57218	0.56185
Ircost	-0.17923	-0.15161	-0.16548	-0.25558	-0.185	-0.20897	-0.29987	-0.25784
Turnover	2.2979	1.9121	2.0462	2.0924	1.8823	1.8899	2.1556	2.5039

Rankcov

Rankcov	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.0785%	0.0784%	0.0782%	0.0776%	0.0777%	0.0776%	0.0774%	0.0767%
STD	1.9419%	1.9526%	1.9684%	1.9904%	2.0181%	2.0511%	2.0891%	2.1787%
IR	0.29142	0.28956	0.2864	0.2813	0.27756	0.2728	0.26711	0.25378
VaR(0.05)	3.2689%	3.2451%	3.2213%	0.031976	0.031738	0.032606	0.033105	0.035087
Var(0.01)	5.4019%	5.5938%	5.6672%	0.057406	0.056037	0.055647	0.055354	0.060141
CVaR(0.05)	4.3966%	4.4421%	4.4901%	0.045404	0.045999	0.046685	0.047392	0.049317
Cvar(0.01)	6.1190%	6.1013%	6.0836%	0.061002	0.062023	0.063558	0.065113	0.069297
MaxDD	-7.4008%	-6.8993%	-6.5122%	-0.06538	-0.069166	-0.072953	-0.07674	-0.084314
Cret	1.347	1.3451	1.3415	1.335	1.3315	1.3264	1.3198	1.3025
Cret_cost	0.89666	0.88862	0.8771	0.86108	0.84474	0.82579	0.80468	0.75764
Ircost	-0.011553	-0.017402	-0.025737	-0.037303	-0.04856	-0.061429	-0.075564	-0.10616
Turnover	1.6315	1.6619	1.7034	1.7578	1.824	1.8996	1.9832	2.1717

FAST-MCD

FAST-MCD (95%)	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.1009%	0.0958%	0.0889%	0.0826%	0.0766%	0.0717%	0.0671%	0.0592%
STD	2.0186%	2.0246%	2.0353%	2.0541%	2.0813%	2.1157%	2.1577%	2.2605%
IR	0.36046	0.34137	0.31505	0.28982	0.26552	0.24445	0.22411	0.18888
VaR(0.05)	3.0392%	3.1038%	3.1688%	0.032634	0.033672	0.034833	0.036039	0.038068
Var(0.01)	6.2728%	6.3312%	6.2017%	0.060723	0.059428	0.058519	0.057663	0.062212
CVaR(0.05)	4.5590%	4.5810%	4.6224%	0.046744	0.047577	0.048515	0.049465	0.051754
Cvar(0.01)	7.0494%	7.0705%	7.0872%	0.071421	0.072505	0.073667	0.07484	0.078437
MaxDD	-8.6349%	-8.5723%	-8.5097%	-0.084472	-0.083846	-0.08322	-0.083396	-0.092577
Cret	1.4951	1.4569	1.4058	1.3592	1.3158	1.2792	1.2441	1.1826
Cret_cost	0.91469	0.88751	0.85068	0.81528	0.77973	0.74624	0.71221	0.64679
Ircost	0.0084678	-0.012767	-0.042197	-0.070542	-0.09878	-0.12487	-0.15087	-0.19872
Turnover	1.971	1.9881	2.015	2.0501	2.0988	2.1617	2.2372	2.4203

QC-IQD

QCIQR	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.1378%	0.1358%	0.1331%	0.1300%	0.1271%	0.1238%	0.1202%	0.1130%
STD	1.9609%	1.9592%	1.9639%	1.9749%	1.9916%	2.0135%	2.0406%	2.1092%
IR	0.50682	0.4999	0.48882	0.4746	0.4602	0.4432	0.42481	0.38643
VaR(0.05)	3.1629%	3.1001%	3.1705%	0.031428	0.031276	0.031722	0.03243	0.033658
Var(0.01)	5.5471%	5.7020%	5.8568%	0.060117	0.061666	0.060874	0.059625	0.057726
CVaR(0.05)	4.3569%	4.3716%	4.3919%	0.044119	0.044417	0.044895	0.04549	0.046877
Cvar(0.01)	6.8110%	6.8296%	6.8471%	0.068646	0.06882	0.068995	0.069169	0.069638
MaxDD	-9.8425%	-9.5661%	-9.2897%	-0.090133	-0.087369	-0.084606	-0.081842	-0.079257
Cret	1.8083	1.7906	1.7659	1.7364	1.7089	1.6769	1.643	1.5739
Cret_cost	1.6314	1.598	1.5546	1.5049	1.4554	1.4019	1.3471	1.2388
Ircost	0.4315	0.41658	0.39568	0.37052	0.34441	0.31534	0.28481	0.22289
Turnover	0.41331	0.45693	0.51179	0.57481	0.64457	0.71903	0.79723	0.96114

2D-Winsorization

2D- Winsorization	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.1552%	0.1554%	0.1548%	0.1535%	0.1524%	0.1508%	0.1489%	0.1451%
STD	2.0105%	2.0065%	2.0071%	2.0128%	2.0236%	2.0391%	2.0591%	2.1117%
IR	0.55656	0.55833	0.55608	0.54991	0.5431	0.53338	0.52162	0.4955
VaR(0.05)	3.2197%	3.2275%	3.2249%	0.032225	0.032117	0.032622	0.033759	0.035508
Var(0.01)	5.4144%	5.5172%	5.5693%	0.055731	0.054402	0.053072	0.052179	0.054942
CVaR(0.05)	4.3672%	4.3824%	4.3978%	0.04417	0.044431	0.044822	0.045277	0.046305
Cvar(0.01)	6.4266%	6.4478%	6.4586%	0.064532	0.064478	0.064425	0.064458	0.065435
MaxDD	-9.6585%	-9.4040%	-9.1494%	-0.088949	-0.086403	-0.083858	-0.081313	-0.076222
Cret	1.9626	1.9652	1.9594	1.9457	1.9331	1.9149	1.8931	1.847
Cret_cost	1.8178	1.803	1.7772	1.7423	1.707	1.6657	1.621	1.5303
Ircost	0.50181	0.49669	0.4863	0.47115	0.45479	0.43512	0.41327	0.36729
Turnover	0.30764	0.34582	0.39158	0.44308	0.4993	0.55942	0.62263	0.75479

2D-Huber

2D-Huber	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.1572%	0.1561%	0.1547%	0.1528%	0.1505%	0.1478%	0.1446%	0.1382%
STD	1.9509%	1.9478%	1.9504%	1.9589%	1.9730%	1.9918%	2.0159%	2.0775%
IR	0.5809	0.57777	0.57212	0.56236	0.55001	0.53521	0.51739	0.4796
VaR(0.05)	3.1006%	3.0961%	3.0635%	0.031475	0.032315	0.033155	0.033445	0.034556
Var(0.01)	5.5797%	5.8005%	6.0213%	0.062187	0.061196	0.060204	0.059213	0.057231
CVaR(0.05)	4.3600%	4.3693%	4.3862%	0.044095	0.044343	0.044747	0.045226	0.046649
Cvar(0.01)	6.7600%	6.7730%	6.7861%	0.067991	0.068121	0.068251	0.068382	0.068642
MaxDD	-9.6841%	-9.3878%	-9.0915%	-0.087952	-0.084989	-0.082026	-0.079063	-0.073462
Cret	1.9935	1.9832	1.9697	1.9487	1.924	1.8952	1.8607	1.7903
Cret_cost	1.8268	1.8013	1.7696	1.7289	1.6834	1.6335	1.5785	1.4688
Ircost	0.51662	0.50687	0.49329	0.47465	0.45278	0.42804	0.40013	0.34251
Turnover	0.35028	0.38605	0.42997	0.48035	0.53621	0.59636	0.65995	0.79425

Appendix 4. Performance of V, CAPM, Principal, Mahalanobis, Frobenius, CCC-GARCH, DCC-GARCH models

V1_cross

V1_cross	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.1913%	0.1985%	0.2071%	0.2148%	0.2185%	0.2201%	0.2188%	0.2157%
STD	2.4063%	2.4309%	2.4648%	2.5088%	2.5600%	2.6240%	2.7004%	2.8788%
IR	0.57341	0.58878	0.60581	0.61745	0.61549	0.60482	0.58432	0.5402
VaR(0.05)	3.7266%	3.7105%	3.6683%	0.036427	0.036989	3.8312%	4.0116%	4.4199%
Var(0.01)	6.6854%	6.6496%	6.5384%	0.065083	0.066684	6.8285%	7.1128%	7.3089%
CVaR(0.05)	5.0229%	5.0512%	5.0853%	0.051192	0.051626	5.2437%	5.3820%	5.7617%
Cvar(0.01)	7.2526%	7.3051%	7.3576%	0.074263	0.075309	7.6309%	7.7557%	8.0822%
MaxDD	-8.5504%	-8.3489%	-8.1475%	-0.07946	-0.078497	-8.0536%	-8.2630%	-8.6818%
Cret	2.2528	2.328	2.4205	2.5028	2.5334	2.533	2.4923	2.3944
Cret_cost	2.1921	2.2517	2.3231	2.3828	2.3885	2.3635	2.2986	2.1538
Ircost	0.55732	0.56936	0.58225	0.58976	0.58289	0.56736	0.54168	0.4876
Turnover	0.11014	0.13457	0.16582	0.19855	0.23807	0.27972	0.32635	0.42674

V2_cross

V2_cross	10%	15%	20%	25%	30%	35%	40%	50%
mean	0.1383%	0.1396%	0.1424%	0.1454%	0.1468%	0.1474%	0.1480%	0.1462%
STD	2.0241%	2.0210%	2.0225%	2.0296%	2.0422%	2.0622%	2.0911%	2.1707%
IR	0.49276	0.49817	0.50763	0.51646	0.51831	0.51556	0.51049	0.48571
VaR(0.05)	3.2889%	3.2376%	3.2207%	0.031528	0.032207	3.2434%	3.2558%	3.3590%
Var(0.01)	5.8370%	5.8810%	5.9736%	0.060796	0.061855	6.1464%	6.1049%	6.1894%
CVaR(0.05)	4.4752%	4.4696%	4.4785%	0.044995	0.045339	4.5825%	4.6379%	4.7730%
Cvar(0.01)	6.5602%	6.5619%	6.5733%	0.065874	0.066014	6.6155%	6.6296%	6.6912%
MaxDD	-8.2833%	-8.0939%	-7.9044%	-0.07715	-0.075255	-7.3360%	-7.1466%	-7.1208%
Cret	1.8017	1.814	1.8389	1.8652	1.8762	1.8784	1.8783	1.8456
Cret_cost	1.7538	1.7559	1.7683	1.7802	1.7761	1.7621	1.7456	1.6808
Ircost	0.47371	0.47513	0.47993	0.48362	0.47994	0.47118	0.46026	0.42379

Turnover	0.10829	0.13075	0.1575	0.18752	0.22025	0.25693	0.29441	0.37565

LAD1_cross

LAD1_cross	**10%**	**15%**	**20%**	**25%**	**30%**	**35%**	**40%**	**50%**
mean	0.1926%	0.1993%	0.2053%	0.1853%	0.1777%	0.1865%	0.1822%	0.1638%
STD	2.4055%	2.4301%	2.4636%	2.4207%	2.3678%	2.5203%	2.5910%	2.7446%
IR	0.57751	0.59142	0.60078	0.55186	0.54131	0.53368	0.50713	0.43036
VaR(0.05)	3.7266%	3.7105%	3.6683%	0.036427	0.036642	3.8135%	4.0032%	4.2286%
Var(0.01)	6.6854%	6.6496%	6.5384%	0.065083	0.066496	6.8285%	6.9887%	7.3089%
CVaR(0.05)	5.0192%	5.0446%	5.0773%	0.050914	0.050347	5.2051%	5.3422%	5.6947%
Cvar(0.01)	7.2444%	7.2823%	7.3371%	0.073283	0.072491	7.4665%	7.5726%	7.9451%
MaxDD	-8.5504%	-8.3489%	-8.1475%	-0.077334	-0.082213	-7.9660%	-8.0824%	-8.4096%
Cret	2.2677	2.3378	2.3989	2.1814	2.1139	2.169	2.1042	1.8811
Cret_cost	2.2073	2.2619	2.3031	2.0762	2.0445	2.0231	1.9388	1.6898
Ircost	0.56159	0.57214	0.57735	0.5229	0.52128	0.49433	0.46202	0.37427
Turnover	0.10889	0.13341	0.16458	0.19962	0.13477	0.28096	0.33001	0.432

LAD2_cross

LAD2_cross	**10%**	**15%**	**20%**	**25%**	**30%**	**35%**	**40%**	**50%**
mean	0.1353%	0.1369%	0.1391%	0.1400%	0.1393%	0.1380%	0.1380%	0.1346%
STD	1.9247%	1.9235%	1.9265%	1.9352%	1.9550%	1.9850%	2.0177%	2.1048%
IR	0.50692	0.51317	0.5206	0.52151	0.5138	0.50125	0.493	0.4612
VaR(0.05)	3.2524%	3.2126%	3.2043%	0.031194	0.03087	3.0813%	3.1967%	3.3850%
Var(0.01)	5.7475%	5.8221%	5.8657%	0.059113	0.059196	5.8945%	5.9807%	6.2093%
CVaR(0.05)	4.3023%	4.3045%	4.3246%	0.043561	0.044083	4.4680%	4.5239%	4.6737%
Cvar(0.01)	6.3217%	6.3407%	6.3472%	0.063408	0.063368	6.3454%	6.3433%	6.5625%
MaxDD	-7.8777%	-7.7037%	-7.5400%	-0.074219	-0.072985	-7.1733%	-6.8730%	-6.8805%
Cret	1.7922	1.8067	1.8261	1.8325	1.823	1.8057	1.7995	1.754
Cret_cost	1.7314	1.7364	1.7442	1.7379	1.7153	1.6841	1.6628	1.5885
Ircost	0.48115	0.48355	0.48642	0.48226	0.46917	0.4509	0.43679	0.39351
Turnover	0.13842	0.15917	0.1842	0.21273	0.24445	0.2798	0.31729	0.39768

References

(1) Markowitz, H. (1952) Portfolio Selection. *Journal of Finance 7*, 77-97.
(2) Markowitz, H. (1959) *Portfolio Selection*, 1 ed., Blackwell.
(3) Cremers, J.-H., Kritzman, M., and Page, S. (2003) Portfolio Formation with Higher Moments and Plausible Utility. *Revere Street Working Paper Series*.
(4) Merton, R. (1972) An Analytical Derivation of the Efficient Portfolio Frontier. *Journal of Financial and Quantitative Analysis 7*, 1851-1872.
(5) Sharpe, W. (1964) Capital Asset Prices: A Theory of Market Equilibriums under Conditions of Risk. *The Journal of Finance 19*, 425-442.
(6) Litner, J. (1965) The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics 47*, 13-37.
(7) Dickenso.Jp. (1974) Some Statistical Aspects of Portfolio Analysis. *Statistician 23*, 5-16.
(8) Klein, R. W., and Bawa, V. S. (1976) Effect of Estimation Risk on Optimal Portfolio Choice. *Journal of Financial Economics 3*, 215-231.
(9) Jorion, P. (1985) International Portfolio Diversification with Estimation Risk. *Journal of Business 58*, 259-278.
(10) Sharpe, W. (1963) A simplified Model for Portfolio Analysis. *Management Science 79*, 277-93.
(11) Ledoit, O., and Wolf, M. (2004) Honey, I shrunk the sample covariance matrix - Problems in mean-variance optimization. *Journal of Portfolio Management 30*, 110-+.
(12) Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307-327.
(13) Bollerslev, T., Engle, R., and Wooldridge, J. M. (1988) A Capital Asset Pricing Model with Time-Varying Covariances. *Journal of Political Economy. 96*, 116-131.
(14) Engle, R. (1999), Working paper.
(15) Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992) ARCH Modelling in Finance. *Journal of Econometrics 52*, 5-59.
(16) Bollerslev, T. (1990) Modeling the Coherence in Short-Run Nominal Exchange-Rates - a Multivariate Generalized Arch Model. *Review of Economics and Statistics 72*, 498-505.
(17) Tse, Y. K., and Tsui, A. K. C. (2002) A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics 20*, 351-362.
(18) Tse, Y. K. (2000) A test for constant correlations in a multivariate GARCH model. *Journal of Econometrics 98*, 107-127.
(19) Connor, G., and Korajczyk, R. A. (1993) A Test for the Number of Factors in an Approximate Factor Model. *Journal of Finance 48*, 1263-1291.
(20) Engle, R. F., and Sheppard, K. (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. *working paper*.
(21) Michaud, R. (1989) The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal 45*, 31-42.
(22) Moronna, R. A. (1976) Robust M-estimators of multivariate location and scatter. *The Annals of Statistics 4*, 51-67.
(23) Brunelli, R., and Messelodi, S. (1995) Robust Estimation of Correlation with Applications to Computer Vision. *Pattern Recognition 28*, 833-841.

(24) Dueck, A., and Lohr, S. (2005) Robust estimation of multivariate covariance components. *Biometrics 61*, 162-169.
(25) Eldar, Y. C., and Merhav, N. (2004) A competitive minimax approach to robust estimation of random parameters. *Ieee Transactions on Signal Processing 52*, 1931-1946.
(26) Knez, P. J., and Ready, M. J. (1997) On the robustness of size and book-to-market in cross-sectional regressions. *Journal of Finance 52*, 1355-1382.
(27) Martin, R. D., and Simin, T. T. (2003) Outlier-resistant estimates of beta. *Financial Analysts Journal 59*, 56-69.
(28) Sharpe, W. (1971) Mean-Absolute-Deviation Characteristic Lines for Securities and Portfolios. *Management Science 18*, B1-B13.
(29) Lauprête, G. J. (2001) Portfolio risk minimization under departures from normality. *MIT PhD Thesis*.
(30) Huber, P. J. (1964) Robust estimation of a location parameter. *Annals of Mathematical Statistics 35*, 73-101.
(31) Campbell, N. A. (1980) Robust Procedures in Multivariate Analysis I: Robust Covariance Estimation. *Applied Statistics 29*, 231-237.
(32) Campbell, N. A. (1982) Robust Procedures in Multivariate-Analysis .2. Robust Canonical Variate Analysis. *Applied Statistics-Journal of the Royal Statistical Society Series C 31*, 1-8.
(33) Balvers, R. J., Cosimano, T. F., and Mcdonald, B. (1990) Predicting Stock Returns in an Efficient Market. *Journal of Finance 45*, 1109-1128.
(34) Rousseeuw, P. J. (1984) Least Median of Squares Regression. *Journal of the American Statistical Association 79*, 871-880.
(35) Coakley, C. W., and Hettmansperger, T. P. (1993) A Bounded Influence, High Breakdown, Efficient Regression Estimator. *Journal of the American Statistical Association 88*, 872-880.
(36) Davies, L. (1992) The Asymptotics of Rousseeuw Minimum Volume Ellipsoid Estimator. *Annals of Statistics 20*, 1828-1843.
(37) Rousseeuw, P. J., and Van Driessen, K. (1999) A fast algorithm for the minimum covariance determinant estimator. *Technometrics 41*, 212-223.
(38) Maronna, R. A., and Zamar, R. H. (2002) Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics 44*, 307-317.
(39) Alqallaf:, F. A., Konis, K. P., Martin, R. D., and Zamar, R. H. (2002) Scalable robust covariance and correlation estimates for data mining. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining Statistical methods*, 14 - 23
(40) Khan, J. A., Aelst, S. V., and Zamar, R. H. (2005) Robust linear model selection based on least angle regression. *Working Paper*.
(41) Chilson, J., Ng, R., Wagner, A., and Zamar, R. (2004) Parallel computation of high dimensional robust correlation and covariance matrices. *Proceedings of the ACM SIGKDD*, 533-538.
(42) Jorion, P. (1986) Bayes-Stein Estimation for Portfolio Analysis. *Journal of Financial and Quantitative Analysis 21*, 279-292.
(43) Jorion, P. (1991) Bayesian and Capm Estimators of the Means - Implications for Portfolio Selection. *Journal of Banking & Finance 15*, 717-727.

(44) MacKinlay, A. C., and Pastor, J. (2000) Asset pricing models: Implications for expected returns and portfolio selection. *Review of Financial Studies 13*, 883-916.

(45) Pastor, L. (2000) Portfolio selection and asset pricing models. *Journal of Finance 55*, 179-223.

(46) Black, F., and Litterman, R. (1990) Asset Allocation: Combining Investor Views with Market Equilibrium. *Goldman, Sachs & Co, Fixed Income Research.*

(47) Black, F., and Litterman, R. (1992) Global Portfolio Optimization. *Financial Analysts Journal 48*, 28-43.

(48) Bevan, A., and Winkelmann, K. (1998) Using the Black-Litterman Global Asset Allocation Model:. Three Years of Practical Experience. *Goldman, Sachs & Co. Fixed Income Research.*

(49) He, G., and Litterman, R. (1999) The Intuition Behind Black-Litterman Model Portfolios. *Goldman, Sachs & Co. Investment Management Research.*

(50) (2003) *Modern Investment Management: an Equilibrium Approach*, Wiley.

(51) Idzorek, T. M. (2004) A Step-by-step Guide to the Black-Litterman Model Incorporating user-specified confidence levels. *Zephyr Allocation Advisor.*